

UNIVERSITY OF CHICAGO

PROTEIN FOLDING: NEW METHODS UNVEIL RATE-LIMITING STRUCTURES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY
BRYAN ANDREW KRANTZ

CHICAGO, ILLINOIS

AUGUST 2002

Copyright © 2002 by Bryan Andrew Krantz
All rights reserved

TO MY WIFE, KRIS, AND MY FAMILY

Table of Contents

	<i>Page</i>
<i>List of Figures</i>	viii
<i>List of Tables</i>	x
<i>List of Abbreviations</i>	xi
<i>Acknowledgements</i>	xii
<i>Abstract</i>	xiii
 <i>Chapter</i>	
1.0 Introduction to Protein Folding	1
1.1 Beginnings: Anfinsen and Levinthal	1
1.1.1 Intermediates populated by proline isomerization	2
1.1.2 Intermediates populated by disulfide bond formation	4
1.1.3 H/D labeling folding intermediates	7
1.1.4 Kinetically two-state folding	8
1.2 Present thesis within historical context	11
1.3 The time scales	12
1.3.1 From microseconds to decades	12
1.3.2 A diffusive process	13
1.3.3 Misfolding and chaperones	13
1.4 The energies	14
1.4.1 The hydrophobic effect	14
1.4.2 Conformational entropy	15
1.4.3 Hydrogen bonds	15
1.4.4 Electrostatics	16
1.4.5 Van der Waals	19
1.5 Surface burial	19
1.6 Modeling protein folding	20
1.6.1 Diffusion-collision and hydrophobic collapse	21
1.6.2 Search nucleation	22
1.6.3 Are there intermediates?	23
1.6.4 Pathways or landscapes?	27
 2.0 The Initial Barrier Hypothesis in Protein Folding	 30
2.1 Abstract	30
2.2 Introduction	30
2.3 Early intermediates	41
2.3.1 Cytochrome <i>c</i>	41
2.3.2 Barnase	46
2.3.3 Ubiquitin	50
2.3.3.1 Single-jump measurements	52

	2.3.3.2	Double-jump measurements	60
	2.3.3.3	Continuous-flow measurements	60
	2.3.3.4	Stabilizing additives	65
	2.3.3.5	Importance of stopped-flow dead time	68
	2.3.3.6	Discussion of ubiquitin	68
	2.3.4	Protein G, B1 domain	72
	2.3.5	IM7	77
	2.3.6	CD2.D1	92
	2.3.7	Summary of early intermediates	92
2.4		Late Intermediates	93
	2.4.1	Cytochrome <i>c</i>	93
	2.4.2	Barstar	96
	2.4.3	Summary of later intermediates	96
2.5		Discussion	96
	2.5.1	Stable intermediates	98
	2.5.2	Late barriers	99
2.6		Conclusion	99
2.7		Methods	100
	2.7.1	Proteins	100
	2.7.2	Equilibrium measurements	101
	2.7.3	Kinetic measurements	101
	2.7.4	Data analysis	103
3.0		Protein Hydrogen Bond Network Formation and Backbone Amide Isotope Effects	104
3.1		Abstract	104
3.2		Introduction	105
3.3		Equilibrium isotope effects	113
	3.3.1	Experimental design	116
	3.3.2	Coiled coil results	119
	3.3.3	The time dependence of hydrogen exchange	129
	3.3.4	The correlation with helical proteins	129
	3.3.5	Solvent isotope effects	130
3.4		Estimates of H-bond strength from isotope effects	133
	3.4.1	Fractionation factors and isotope effects	133
	3.4.2	A scale factor for H-bond strength	135
	3.4.3	α -helix versus β -sheet	136
	3.4.4	Limitations and implications of the scale factor result	138
3.5		Kinetic isotope effects	139
3.6		Analysis of kinetic isotope effects	142
3.7		Assessing specific protein folding models with $\phi_f^{\text{D-H}}$	148
	3.7.1	GCN4 coiled coil	148
	3.7.2	Cytochrome <i>c</i>	148
	3.7.3	Arc repressor	149
	3.7.4	Lambda repressor	151
	3.7.5	IM7	151

3.7.6	CTACP	152
3.7.7	CI2	153
3.7.8	Ubiquitin	154
3.7.9	Protein G	154
3.7.10	Protein L and SH3	155
3.7.11	CD2 and hen egg white lysozyme	155
3.8	Transition state structure and general folding models	156
3.9	Conclusions	162
3.10	Methods	162
3.10.1	Proteins	162
3.10.2	Equilibrium isotope effect measurements	163
3.10.3	Stopped-flow kinetics	163
3.10.4	Data analysis	164
4.0	Engineered Divalent Metal Ion Binding Sites and Pathway Heterogeneity	165
4.1	Abstract	165
4.2	Introduction	166
4.3	The coiled coil, GCN4	169
4.3.1	Equilibrium binding stability of designed binding sites	169
4.3.2	Cross-linked coiled coil and pathway homogeneity	177
4.3.3	Quantifying heterogeneity: ψ -analysis	180
4.3.4	Pathway heterogeneity in dimeric GCN4	184
4.3.5	Properties of the coiled coil transition state	189
4.3.6	Metal binding and ψ -values	190
4.3.7	Conclusions	190
4.4	Ubiquitin	191
4.4.1	The α -helical sites	191
4.4.2	The β -sheet sites	196
4.4.3	Secondary structure metal ion preferences	199
4.4.4	Conclusions for ubiquitin	200
4.5	Other studies of pathway heterogeneity	203
4.6	Methods	204
4.6.1	Proteins	204
4.6.2	Equilibrium	205
4.6.3	Stopped-flow kinetics	205
5.0	Conclusions and Future Directions	206
5.1	General conclusions	206
5.2	Future directions	209
<i>Appendices</i>		
A1	Protein Expression Vectors and Mutants	210
A2	Protein Sequences and Purification Methods	213

A2.1	General protein expression and purification methods	213
A2.1.1	Expression cell lines and competency protocols	213
A2.1.2	Protein expression and lysis	214
A2.1.3	Acetic acid precipitation	217
A2.1.4	Ion exchange chromatography	217
A2.1.5	Six-his affinity chromatography	219
A2.1.6	Hydrophobic interaction chromatography	219
A2.1.7	HPLC	220
A2.1.8	Dialysis	221
A2.1.9	Freeze-drying and storage	221
A2.1.10	Reconstitution of freeze-dried proteins	221
A3	Two-State Equilibrium and Kinetic Analysis	227
A3.1	Monomeric systems	227
A3.1.1	Two-state equilibrium stability derivation	227
A3.1.2	Two-state kinetics derivation	228
A3.2	Dimeric systems	229
A3.2.1	Two-state equilibrium stability derivation	229
A3.2.2	Two-state kinetics equations	230
A3.3	ϕ -analysis	231
A4	Optimizing the Stopped-flow	235
A4.1	Hardware	235
A4.1.1	Redesigned observation head	235
A4.1.2	Redesigned hard stop case	236
A4.1.3	Optics	241
A4.1.3.1	Mercury versus xenon arc lamps	241
A4.1.3.2	Reducing observation head background light scattering	242
A4.1.3.3	Fluorescence versus absorbance measurements	242
A4.1.4	Electronics	242
A4.1.4.1	Hamamatsu photomultiplier tube	242
A4.1.4.2	Electronic interface to A/D board	243
A4.1.4.3	Signal filtering and over-sampling	243
A4.2	Software	258
A4.2.1	Post-acquisition logarithmic filtering	258
A4.2.2	Microcal Origin software scripts	258
Works Cited		262

List of Figures

<i>Figure</i>	<i>Page</i>
1.1 Typical depictions of protein folding reaction coordinates	17
1.2 Putative models for protein folding	24
1.3 Branched landscapes versus linear folding pathways	28
2.1 Two-state versus three-state folding	31
2.2 The burst phase and the chevron rollover in protein folding	34
2.3 Cyt <i>c</i> folding	39
2.4 Barnase folding and unfolding	47
2.5 Ubiquitin folding traces	53
2.6 Ubiquitin denaturant dependence	57
2.7 Ubiquitin continuous-flow measurements	62
2.8 Ubiquitin folding in the presence of a stabilizing additive	66
2.9 Dead time and the accurate determination of the folding rate of ubiquitin	69
2.10 Mono-exponential two-state folding of Protein G	73
2.11 Direct comparison of Protein G results to Park <i>et al.</i>	75
2.12 Protein concentration dependence of IM7 refolding in GdmCl	79
2.13 Protein concentration dependence of IM7's burst phase in urea	81
2.14 The denaturant dependence of IM7 refolding on rates and amplitudes	83
2.15 Continuous-flow SAXS on IM7	86
2.16 Stopped-flow CD studies on IM7 refolding	90
3.1 Three-dimensional renderings of model proteins	106
3.2 Hydrogen bond composition for small two-state model proteins	111
3.3 A thermodynamic cycle for amide isotope effects	114
3.4 Experimental design of the equilibrium isotope effect measurements	117
3.5 Equilibrium isotope effect time courses for α -helical proteins	120
3.6 Equilibrium isotope effect time courses for β -sheet and α/β proteins	122
3.7 Denaturant dependence of the equilibrium isotope effects	124
3.8 Equilibrium isotope effect correlates with helical H-bond composition	131
3.9 Four syringe protocol for a Biologic SFM4 stopped flow	140
3.10 Folding kinetic isotope effect measurements for α -helical proteins	144
3.11 Folding kinetic isotope effect measurements for mixed α/β and β -sheet proteins	146
3.12 H-bond formation and surface burial are concomitant folding processes	158
4.1 Designed metal ion binding variants of the coiled coil	167
4.2 Equilibrium stability obtained by the addition of the divalent metal ions	170
4.3 Denaturant dependence of folding in presence of metal ions for cross-linked GCN4	174
4.4 Determining folding pathway models for cross-linked GCN4	178
4.5 Denaturant dependence of folding in presence of metal ions for dimeric GCN4	182
4.6 Determining folding pathway models for dimeric GCN4	185
4.7 Three-dimensional rendering of designed metal binding sites in ubiquitin	192
4.8 ψ -analysis of α -helical sites in ubiquitin	194
4.9 ψ -analysis of a β -sheet site in ubiquitin	197
A3.1 ϕ -values and chevron plots	233

<i>Figure</i>		<i>Page</i>
A4.1	Redesigned stopped-flow observation head	237
A4.2	Redesigned hard stop with heat sink and fan cooling	239
A4.3	Overall layout of the Biologic stopped-flow instrument	244
A4.4	PMT power supply	246
A4.5	PMT control and signal filtering electronics	248
A4.6	PMT signal pre-filtering with an LRC	250
A4.7	Signal quality of low-pass electronic filters	255

List of Tables

<i>Table</i>		<i>Page</i>
2.1	Thermodynamic and kinetic parameters for ubiquitin folding	55
2.2	Stopped and continuous-flow measurements for ubiquitin	64
3.1	Model systems for backbone amide isotope effect studies	108
3.2	Equilibrium and kinetic backbone amide isotope effects parameters	126
4.1	Equilibrium metal ion stabilization of biHis GCN4 variants	173
4.2	Kinetic parameters for metal ion binding to biHis GCN4 variants	176
4.3	Heterogeneous and homogeneous folding model fit parameters	187
4.4	Chevron analysis of amino acid substitutions for Ub metal site variants	201
4.5	Kinetic folding pathway parameters from ψ -analysis for ubiquitin	202
A1.1	Expression vectors from total DNA synthesis	211
A1.2	Expression vectors and mutants	212
A2.1	Purification schemes	222
A2.2	Protein sequences	223
A4.1	Pinouts for cables created to interface the electronic components	252
A4.2	Complete parts list for stopped-flow modifications	253
A4.3	Quantitative comparison of low-pass electronic signal filters	257
A4.4	Dataset names used by the Origin scripts	258
A4.5	Origin scripts	259

List of Abbreviations

Arc	arc repressor
Arc ^{MYL}	arc repressor, R31M/E36Y/R40L
biHis	designed two histidine site for divalent metal ion binding
BPTI	bovine pancreatic trypsin inhibitor
CI2	chymotrypsin inhibitor 2 from barley
CTACP	common type acylphosphatase
Cyt <i>c</i>	cytochrome <i>c</i>
Cyt <i>c</i> ^{pWT}	cytochrome <i>c</i> pseudo wild type
D	deuterium
EDTA	ethylenediaminetetra acetic acid
Fl	fluorescence
GCN4 ^{Cx}	GCN4 coiled coil p1' domain cross-linked at C-terminus
GCN4 ^d	dimeric GCN4 coiled coil p1' domain
GCN4 ^{Nx}	GCN4 coiled coil p1' domain cross-linked at N-terminus
GdmCl	guanidinium chloride
H-bond	hydrogen bond
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2ethanesulfonic acid)
HPLC	high performance liquid chromatography
HX	hydrogen exchange
IM7	colicin immunity protein 7
IPTG	isopropylthiogalactoside
λ ^{AA}	monomeric lambda repressor, residues 6-85, G46A/G48A
λ ^{GG}	monomeric lambda repressor, residues 6-85
M ²⁺	divalent metal ion
NMR	nuclear magnetic resonance
PMT	photomultiplier tube
ND protein	ND designates protein's backbone amides are deuterated
^{NH} protein	NH designates protein's backbone amides are protonated
ProtG	protein G, B1 domain
ProtL	protein L, B1 domain
RNase	ribonuclease
SAXS	small angle x-ray scattering
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	<i>src</i> -SH3 domain
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TRIS-Cl	2-amino-2-hydroxymethyl-1,3-propanediol hydrochloride
TS	transition state
Ub	mammalian ubiquitin with F45W substitution

Acknowledgements

I am extremely grateful for the successes I have been able to enjoy while receiving my doctoral training at the University of Chicago. These successes are only possible due to the people who have mentored and trained me over the last five years. The Department of Biochemistry and Molecular Biology has become a vibrant and stimulating environment to do science, and I thank those persons that continue to shape and improve the department, including the BMB office staff, Mike Clark (general manager), Lisa Thompson (graduate student shepherd), and Tony Kossiakoff (CEO). Of course, the great science I was able to partake in was only available due to the stimulating environment of Tobin Sosnick's lab. In Tobin's lab, I was able to not only benefit from countless close interactions with the boss but also with some talented post-docs, who include Xingwang Fang and Liam Moran, and other graduate students in the group. Meetings with my thesis committee, Tao Pan (chair), Susan Lindquist, Steven Kent, and Tony Kossiakoff, afforded numerous thought provoking discussions as well as criticisms. Tao Pan has been not only a great chairperson of my thesis committee but also a great general advisor, keeping me focused on important scientific questions, finding a post-doctoral position and, of course, finishing my thesis. Walter Englander mentored me, providing perspective from outside of the University of Chicago, as well as aided in my post-doctoral search. I would also like to thank collaborators who contributed research materials. Bob Sauer at MIT contributed purified Arc repressor proteins. David Baker and Keith D. Wilkinson provided plasmids capable of expressing recombinant model proteins, Protein G, Protein L, and ubiquitin. Finally, I wish to thank my family for hanging in there and supporting my decision to do scientific research, and, of course, my wife, Kris, for her unwavering support of my career.

Abstract

Protein folding is the final step in the Central Dogma of biology whereby newly translated proteins achieve their unique three-dimensional structures. With the goal of determining the general biophysical laws that govern the process of protein folding, I rely on traditional methods while developing several novel experimental techniques.

First, the rule of two-state folding for small globular protein is investigated by testing known exceptions to this rule. The exceptions to the phenomenon are often mistakenly due to protein aggregation, misfolding, solvent dependent relaxation, or multiple mismatched data sets. Therefore, the Initial Barrier Hypothesis is maintained for nearly all small globular proteins, which asserts that the major barrier to protein folding is limited by a search process and intermediates more stable than the unfolded state do not appear prior to this barrier.

Second, a novel isotope effect measurement of backbone amide hydrogen bonds is developed. An average hydrogen/deuterium equilibrium isotope effect of 8.6 cal/mol/bond is assessed only for bonds in helical geometries. The number of helical hydrogen bonds that form in the transition state ensemble is determined using the kinetic partitioning of the isotope effect. Hydrogen bonds and solvent accessible surface area are buried concomitantly. It is hypothesized that the condensation of hydrophobic side chains excludes water from hydrogen bond partners, subsequently driving their formation.

Third, a novel method, whereby bi-histidine metal ion binding sites are engineered into proteins, assesses whether multiple folding pathways exist by implementing ψ -analysis. ψ -analysis resolves the limitations of traditional mutational ϕ -analysis in that both the proportion of molecules traversing a given pathway and fractional formation of a given site are assessed. Fold topology is found to play a crucial role in governing the degree of pathway heterogeneity for a protein.

1

Introduction to Protein Folding

1.1 Beginnings: Anfinsen and Levinthal

The process by which a newly translated and unfolded polypeptide chain achieves its native conformation is of profound importance to the biology of the cell. Generally, the native structure for a particular polypeptide is unique, and the folded structure defines the protein's function and role in the cell. Anfinsen demonstrated that a functional enzyme such as ribonuclease could be denatured, disrupting catalytic activity, and that the denatured enzyme could then be renatured and reactivated by simply diluting away the denaturant and re-oxidizing broken disulfide bonds (Anfinsen, 1973). The primary amino acid sequence, therefore, contains all the necessary information required for a protein to find its native conformation. Furthermore, the folding of a protein is generally a spontaneous process that requires no additional protein factors or expenditure of ATP.

Since Anfinsen's remarkable demonstration, many other small proteins have been unfolded and refolded, and many of the smaller single domain proteins (~110 amino acids in length) fold in less than a second or even a millisecond (Burton *et al.*, 1996; Crane *et al.*, 2000; Durr *et al.*, 1999; Huang & Oas, 1995; Jager *et al.*, 2001; Moran *et al.*, 1999; Sosnick *et al.*, 1996b; Sosnick *et al.*, 1994; Sosnick *et al.*, 1995). The latter fact invokes Levinthal's paradox, stating that due to the vast number of potential conformations a polypeptide can traverse, it should not fold in a tangible amount of time

if the polypeptide searches randomly (Levinthal, 1968; Levinthal, 1969). To resolve the paradox, proteins traverse a limited subset of conformations or fold via a folding pathway (Levinthal, 1969).

The earliest folding studies, therefore, sought to “trap” folding intermediates and identify the regions of the protein, which became structured in the intermediate (Kim & Baldwin, 1982; Kim & Baldwin, 1990). Generally, these studies took advantage of *cis/trans*-proline isomerization in ribonuclease A (RNase A), disulfide bond shuffling in bovine pancreatic trypsin inhibitor (BPTI), and histidine-heme mis-ligation in cytochrome *c* (Cyt *c*) to trap intermediates for further characterization.

1.1.1 Intermediates populated by proline isomerization

Baldwin and co-workers identified fast and slow unfolded forms of RNase A, U_F and U_S , respectively (Garel & Baldwin, 1973). These unfolded species were proposed to interconvert slowly leading to the biphasic kinetics observed. Proline isomerization was cited by Brandts *et al.* to explain the three-state mechanism (Eq. 1.1; Brandts *et al.*, 1975)



The populations of U_F and U_S , were estimated at 20% and 80%, respectively, where initial conditions, such as denaturant and temperature, did not affect the unfolded state distribution (Garel & Baldwin, 1973; Garel *et al.*, 1976). Because two specific sites require the *cis*-proline isomer in native RNase A, only 20% of the unfolded protein occurs with the correct *cis*-prolines as the U_F species.

Experiments which concluded that *cis/trans* proline isomerization was operant included comparison to model compounds by Grathwohl and Wuthrich as well as the demonstration that the unfolded populations interconverted with a high activation enthalpy of 20 kcal mol⁻¹ (Brandts *et al.*, 1975; Grathwohl & Wuthrich, 1976a; Grathwohl & Wuthrich, 1976b). The residue proceeding the proline most significantly affects the rate of isomerization (Grathwohl & Wuthrich, 1976a; Grathwohl & Wuthrich, 1976b; Wuthrich & Grathwohl, 1974). Direct evidence that the unfolded populations contained *cis/trans*-proline isomers was provided by enzymatically supplementing the folding reaction with porcine kidney *cis/trans*-proline isomerase (Fischer & Bang, 1985). However, while this enzyme eliminated the slow folding phases of many proteins (Lang *et al.*, 1987), it failed to affect the major slow phase of RNase A (Fischer & Bang, 1985; Lang *et al.*, 1987). Site directed mutagenesis definitively demonstrated that proline residues were required for the slowly equilibrating unfolded populations in yeast iso-1 (Ramdas & Nall, 1986) and iso-2 (White *et al.*, 1987) Cyt *c*, thioredoxin (Kelley & Richards, 1987), and RNase T1 (Kiefhaber *et al.*, 1990).

Most other proteins containing proline do not populate intermediates as dramatically as RNase A, since at equilibrium in the unfolded state, *trans*-proline is favored four-to-one over *cis*-proline, and native proline is most often in the *trans* conformation. Moreover, energetic penalties for incorrect proline isomers vary site specifically as predicted from crystal structures, since loop regions that contain proline may accommodate either isomer (Levitt, 1981). Finally, *cis/trans*-proline populated intermediates are not obligate but are rather a consequence of slowly interconverting Xaa-Pro peptide bonds and do not explain what ultimately limits the unencumbered folding reaction, $U_F \leftrightarrow N$.

1.1.2 Intermediates populated by disulfide bond formation

The *in vitro* process of disulfide bond formation (for review Creighton, 1978; Creighton, 1985) was extensively studied by Creighton followed by further characterization by Weissman and Kim (Weissman & Kim, 1991; Weissman & Kim, 1992a; Weissman & Kim, 1992b; Weissman & Kim, 1993; Weissman & Kim, 1995) on the model system, bovine pancreatic trypsin inhibitor (BPTI). BPTI contains three pairs of cysteine residues that form unique disulfide bonds in the native protein. Often these disulfides are referenced as cysteine residue numbers enclosed in brackets, i.e. the native disulfides are [30-51], [5-55], and [14-38]. Because fully reduced protein, in which all these native disulfides are broken, is unfolded, the folding pathway for BPTI is believed to be an excellent model system for trapping intermediates with thiol blocking reagents.

Creighton's pioneering method of studying the folding pathway of BPTI was to trap free reactive thiol groups with iodoacetic acid or iodoacetamide. He demonstrated that about 60% of the earliest single disulfide intermediates were [3-51], 3% were [5-55] or [5-51], and the remainder (30%) were non-native species. He also trapped two major native two disulfide species, [30-51; 14-38] and [5-55; 14-38], as well as lesser amounts of two other two disulfide intermediates which contained a non-native disulfide, [30-51; 5-14] and [30-51; 5-38]. He interpreted the results as follows. The first major single disulfide species [30-51] populates. Then the pathway bifurcates to either (a) double disulfide intermediates that are *completely native*, [5-55; 14-38] and [30-51; 14-38], or (b) those that contain *one non-native disulfide*, [30-51; 5-14] and [30-51; 5-38]. He hypothesized that the native two disulfide species were dead-end species, whereas the non-native species were obligate non-native species required for the formation of the

native state (Creighton, 1977; Creighton, 1978). Thus, the native two disulfide species were kinetically trapped and had to reshuffle to the non-native species in order to form the fully native pre-cursor, [30-51; 5-55], which rapidly converts to the full complement of the three native disulfides.

Creighton's method in theory should have prevented further disulfide formation; however, the integrity of the blocking reaction chemistry was questioned by Weissman and Kim. Either the highly charged acid or polar amide blocking compounds from iodoacetic acid and iodoacetamide, respectively, may not have adequately penetrated into the hydrophobic core and modified all free thiol groups. Creighton's method, therefore, artificially selected for surface exposed thiol groups, favoring the formation of non-native disulfides. This fact was verified by Weissman and Kim, who relied on acid quenching the refolding reaction and purifying the trapped intermediates by modern HPLC techniques (Weissman & Kim, 1991). They found that negligible amounts of non-native disulfides formed at pH 8.7 and almost no non-native species formed at pH 7.3. Moreover, The major one disulfide species was [5-55] at pH 7.3 not [30-51] as both Creighton and later Kim had observed at pH 8.7.

The disulfide bond formation problem, while important to many extra-cellular factors and enzymes, is less interesting to the more general physics of protein secondary and tertiary structure formation. Both Kim and Creighton agree that the formation of disulfides, even in a simple protein, like BPTI, is chemically hampered at physiological pH, since free thiol groups are protonated, potentially buried, and consequently nonreactive. Therefore, the complete formation of all the native disulfides is slow. While Kim argued that the protein required global unfolding (i.e. 6 M urea sped up the process of forming all

three native disulfides), Creighton maintained that BPTI must rearrange through a non-native disulfide species to form all three native disulfides. The putative “obligate” intermediates, in fact, had little proclivity to make the rate-limiting disulfide rearrangement to [30-51; 5-55], invalidating the non-native pathway further (Weissman & Kim, 1992a).

Formation of a single disulfide, [5-55], in an alanine substituted variant (where all other cysteines are replaced with alanine), is identical to native BPTI by 2D NMR spectroscopy (Staley & Kim, 1992). Crystal structures and other NMR structures verified that double disulfide intermediates were also native-like in terms of secondary and tertiary structure (Eigenbrot *et al.*, 1990; Naderi *et al.*, 1991; Stassinopoulou *et al.*, 1984; States *et al.*, 1984). Thus for the sake of argument, BPTI’s folding pathway may be dissected into two different folding problems: (i) formation of the correct overall native topology, secondary and tertiary structure, which include various *non-covalent* side chain hydrophobic interactions, backbone hydrogen bonds, etc., and (ii) formation of all three *covalent* disulfide bonds. Because the structure becomes native in the earliest [5-55] intermediate, and [5-55] is the dominant single disulfide species at pH 7.3, then the non-covalent folding problem is essentially solved at the earliest stage of folding. Furthermore, the non-covalent native-like interactions selectively drive the formation of the [5-55] disulfide bond, since that species dominates the earliest events instead of non-native species or the less stable [30-51] native species. The folding process is complete at this very early single disulfide stage, and thus understanding the rate-limiting step for the formation of the earliest [5-55] species is most crucial to understanding how proteins adopt native secondary and tertiary structures.

Remarkably, the final disulfide reshuffling mechanism made little difference to the rate of creating functional inhibitor. All of the heavily populated native two disulfide intermediates make functional inhibitors. Therefore, from a biological standpoint, driving the formation of secondary and tertiary structure through non-covalent interactions is more crucial to the rapid (millisecond) gain of three-dimensional structure and function, whereas forming the correct complement of disulfides is slow (hours) and may be required to extend BPTI half-life by increasing the protein's equilibrium stability, slowing the unfolding rate and/or increasing the refolding rate and preventing proteolytic degradation.

Neither Kim nor Creighton entertain the possibility that all the intermediates they identify may not in fact be crucial to the actual folding pathway, but rather for BPTI to best solve form disulfides, the best route may be through only a high energy, native, sequential pathway. These intermediates may not populate appreciably, because the later rate constants may be faster than the initial step of forming [5-55]. The method of trapping intermediates may fail more generally in determining the actual folding pathway, since captured species exist only on the slow misfold/reorganization pathways, while the fast folding pathway is seemingly non-existent (as sensed by trapping methods).

1.1.3 H/D labeling folding intermediates

When it became technically feasible to resolve individual protein amide hydrogen with 2D NMR, hydrogen/deuterium exchange (HX) pulse labeling studies of protein folding intermediates were performed by both Englander and Baldwin's groups on Cyt *c* and RNase A, respectively. For Cyt *c*, the trapped intermediate had significantly protected amide hydrogen only in the amino- and carboxy-terminal helices, which were

protected from exchange with solvent in about 10-20 msec, whereas the 60s and 70s helices formed later (~200 msec). The early intermediate of RNase A, that appeared kinetically in the fast phase, contained protected amides throughout the β -sheet network. The tertiary structure of the intermediate, probed by tyrosine absorbance, was believed to lag behind the secondary structure formation indicated by the HX protection; however, alternative models, in which secondary and tertiary structure formed in concert or that the intermediate was due to *cis/trans*-proline isomerization, could not be ruled out. Both Baldwin and Englander's groups interpreted their data in terms of sequential folding models in which intermediates of progressively more native-like structure occurred along a linear folding pathway (Fig. 1.3*b*). Notice that these native-like models are in stark contrast to the obligate non-native disulfide model Creighton had put forth for BPTI folding. The HX labeling experiments performed on Cyt *c* and RNase A were seminal and marked the beginning of numerous HX labeling studies of protein folding (for review see Englander, 2000; Englander & Mayne, 1992). The method is powerful, since site resolved millisecond resolution is achieved wherein stabilities of individual intermediates may be determined simultaneously.

1.1.4 Kinetically two-state folding

Quite early on, Lumry, Biltonen and Brandts argued that the reversible process of folding for small proteins was two-state: $U \leftrightarrow N$ (Lumry *et al.*, 1966). However, the current thinking in the field of protein folding during the 1970s and 80s was that proteins folded through populated intermediates which were arranged linearly like a metabolic pathway (for review see Kim & Baldwin, 1990; see also Fig. 1.3*b*). It was believed that

the step-wise formation of structured intermediates "solved" Levinthal's paradox (see Sec. 1.1). Thus, the notion of two-state folding fell out of favor mainly because proline and disulfide bond intermediates accumulated, making the otherwise intractable problem of studying protein folding more accessible experimentally.

In the early to mid-90s, however, the field of protein folding experienced a very significant paradigm shift. First, Jackson and Fersht elegantly demonstrated that chymotrypsin inhibitor II from barley (CI2) folded in a two-state manner, since the equilibrium free energy and surface area burial parameters were recapitulated in the folding kinetic measurements (Jackson & Fersht, 1991a). Thus, stable intermediates did not have to accumulate before the rate-limiting step, else the equality could not be maintained. "Chevron analysis," previously developed by Matthews, gained popularity with Jackson and Fersht's kinetic studies (Matthews, 1987).

Focus, therefore, shifted from looking for methods that could identify and trap accumulated intermediates, (i.e. those described by Baldwin and Creighton; Secs. 1.1.1-2), to understanding the structure of the highest energy, rate-limiting species, called the transition state. (More precisely, the "transition state" is better referred to as a "transition state ensemble," since many different species may be rate-limiting if there are parallel folding routes; see Fig. 1.3a; Sec. 1.6.4 and Ch. 4.) Understanding the elusive transition state (Fig. 1.1b) was believed to be the missing-link that connected the unfolded ensemble to the native state.

Furthering this end of understanding the structure acquired in the rate-limiting species, mutational ϕ -analysis was applied to CI2 (Itzhaki *et al.*, 1995). In such analysis, amino acids are changed, and the kinetic partitioning of the change in equilibrium free

energy was used to indicate the degree to which the side chain participated in the rate-limiting structure (or structures). While many mutations yielded *partial* ϕ -values, Fersht concluded that partial ϕ -values indicate that the side chain's interactions were *partially* formed in the rate-limiting step. Alternative multiple transition state models were discounted by the clever application of Brønsted analysis (Fersht *et al.*, 1994) although analysis of solvent accessible amino acid substitutions in CI2 showed that pathway heterogeneity remained possible (see App. in Kim *et al.*, 1998).

Contemporary to these studies, Englander's group more carefully examined the multi-phasic folding kinetics of Cyt *c*. It was expected that the protected amide hydrogen would map the precise structure of the folding intermediate of Cyt *c*, and this intermediate was obligate. However, Cyt *c*'s folding intermediate was only stabilized and observable, because mis-ligated histidine-heme complexes caused an error in the folding pathway. Thus, a barrier was inserted in the normal two-state folding pathway (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1994; Sosnick *et al.*, 1995). When the mis-ligation barrier was removed by lowering the buffer pH to 4.8, or adding imidazole or azide to the refolding buffer, the protein folded in an apparent two-state manner according to multiple probes including individually monitored amide hydrogen located throughout the protein. This result was significant, because it demonstrated that the accumulation of intermediates is not obligate. Furthermore, it was then hypothesized that the major barrier to folding is limited by a global search for a native-like nucleus (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995); and thus experiments describing the rate-limiting nucleus were most crucial to understanding how proteins fold (see also Sec. 1.6.2 and Ch. 2).

1.2 Present thesis within historical context

The intellectual history of protein folding studies shows two remarkable changes of emphasis. While Brandts initially hypothesized that proteins might obey all-or-none transitions (or fold in a two-state manner), his insight was generally ignored by Baldwin, Creighton and others. Protein folding was believed to be analogous to a linear, multi-step metabolic pathway with distinctly populated intermediates and not a simple, two-state transition. Thus, research in the 70s and 80s focused on trapping intermediates. Ironically, the intermediates found represented misfolding events (Sosnick *et al.*, 1994) linked to *cis/trans*-proline isomerization or kinetic traps linked to disulfide reshuffling. The focus on intermediates ignored what was rate-limiting when protein folded in an “error free” manner while gaining functional secondary and tertiary structure. For example, the $U_F \leftrightarrow N$ transition for RNase A is the error free folding pathway, which rapidly folds without the accumulation of folding intermediates. The formation of the [5-55] disulfide in BPTI represents the formation of functional inhibitor with identical secondary and tertiary structure to the three disulfide bond containing native form, albeit a less stable form.

Fortunately, during the early to mid-90s the field was refocused by two important works (Jackson & Fersht, 1991a; Sosnick *et al.*, 1994) that revealed that proteins indeed can fold by the very two-state mechanism Brandts and co-workers suggested 30 years earlier. The emphasis of many groups was shifted to understanding what the nature of the high energy rate-limiting species was that bridged the unfolded and native states. The question was then asked: what limits rapid error-free folding when the processes of improper proline isomers, disulfide bond shuffling, protein aggregation, and histidine-heme mis-ligation are ignored? This thesis presents original research on the nature of the

rate-limiting step of two-state folding by first reexamining some apparent exceptions to two-state folding (Ch. 2), second developing a novel method to determine the fraction of hydrogen bonds required in the rate-limiting step (Ch. 3), and third creating an second novel method to assess the diversity of the folding transition state ensemble (Ch. 4). First, however, some additional background material is presented, covering the energies, timescales, and other aspects crucial to understanding the nature of how proteins fold.

1.3 The time scales

1.3.1 From microseconds to decades

Remarkably protein folding time scales cover an extremely broad dynamic range: microseconds to milliseconds for small helical proteins, milliseconds to seconds for largely β -sheet proteins, minutes for proline *cis/trans* isomerization, minutes to hours for repairing misfolded proteins, and decades for prion or amyloid diseases to emerge.

Plaxco and Baker very elegantly noted that the folding rates for two-state proteins vary directly with the average sequence distance between contacting residues in a protein, or contact order (Plaxco *et al.*, 1998). Thus, more local α -helical proteins fold faster than β -sheet proteins, because the average distance between contacting residues is greater in the β -sheet protein which has non-local, distant contacts across β -strands. Many groups have shown similar results (Chiti *et al.*, 1999; Fersht, 2000; Grantcharova *et al.*, 2001; Munoz & Eaton, 1999; Plaxco *et al.*, 2000); other studies (Debe & Goddard, 1999) drew similar conclusions using first principles although this first principles method failed at predicting α -helical protein folding rates.

Experimentally, many of these time scales are readily accessible using NMR coupled to hydrogen exchange (HX) (Bai *et al.*, 1995b) or line broadening measurements (Huang & Oas, 1995), continuous-flow (Pollack *et al.*, 1999; Shastry *et al.*, 1998), or temperature jumps (microsecond to millisecond) (Munoz *et al.*, 1997; Sabelko *et al.*, 1999; Thompson *et al.*, 1997); stopped-flow techniques (millisecond to minutes); or manual dilution (minutes to hours).

1.3.2 *A diffusive process*

As with any process taking place in aqueous water, the speed at which the peptide may fluctuate and sample conformations is limited by diffusion. In fact, the rate-limiting step for many small two-state folding proteins appears diffusion controlled (Bhattacharyya & Sosnick, 1999; Bieri *et al.*, 1999; Creighton, 1997; Jacob *et al.*, 1999; Jacob *et al.*, 1997; Jacob & Schmid, 1999; Plaxco & Baker, 1998; Viguera & Serrano, 1997). This fact reveals that the protein does not experience a significant collapse prior to the rate-limiting step, where only a re-arrangement would be required which would lack a diffusion dependence. The protein is rather limited by the movement of peptide chain through solution, and thus the rate is ultimately the product of the diffusion limit and the average sequence distance among contacting residues and an unknown factor for the success frequency, or productiveness, of given encounters.

1.3.3 *Misfolding and chaperones*

Those proteins which fold slowly, misfold, or aggregate may often require assistance to fold in a manner that is timely to the biologically dictated time scale coinciding with the protein's functional role. Cells have evolved elaborate machinery to assist in folding many of

these proteins. While many proteins fold autonomously, conditions arise when the temperature suddenly increases, driving proteins to misfolded states. Thus, many chaperones are over-expressed during heat shock stress. However, other chaperones are required to assist folding under normal growth conditions. Chaperones function by providing an improved environment for folding, effectively catalyzing the refolding process and reducing the time required to refold. They generally hydrolyze ATP and create a cavity that can accommodate a single polypeptide chain to eliminate intermolecular aggregation (Ellis, 2001; Grantcharova *et al.*, 2001; Hartl & Hayer-Hartl, 2002).

1.4 The energies

Understanding the nature of the energies and forces relevant to protein stability is fundamental to the process of describing protein folding kinetics. This section briefly highlights the energies that guide a protein to its native structure.

1.4.1 The hydrophobic effect

The hydrophobic effect has been described as simply the consequence of water's preference for itself. In order to solvate non-polar hydrocarbon, water is proposed to form strained clathrate networks of hydrogen bonds. The driving force for the hydrophobic effect—the source of the energy—is in these strained hydrogen bonds and the entropy of their organized, ice-like structures. To reduce strain and liberate free energy stored in these clathrate networks, hydrophobic side chains cluster and subsequently bury non-polar surface area (see Tanford, 1980 and references therein). Many results confirm this phenomenon, such as transfer studies of compounds between aqueous and organic solvents. The partitioning equilibrium constants yield transfer free energies, which show that the

non-polar, hydrophobic surface area is directly proportional to the transfer free energy. Finally, energy provided to the protein folding reaction from the hydrophobic effect is probably the most substantial quantity of stabilizing energy; albeit this point is often debated (Pace *et al.*, 1996; see also Sec. 1.5).

1.4.2 Conformational entropy

Opposing the hydrophobic effect in terms of energy are the large numbers of available conformations the protein can possess in the unfolded state. The fact that so many possible states can be sampled also forms the core of the Levinthal paradox. Often conformational entropy is broken down into side chain conformational entropy and backbone conformational entropy. Folding funnels (Bryngelson *et al.*, 1995; Dill & Chan, 1997; Leopold *et al.*, 1992; Onuchic *et al.*, 1996), which are now used ubiquitously to describe the protein folding reaction coordinate, depict the width of the funnel as the broad number of conformational states accessible and the depth of the funnels as the level of enthalpic stabilization (Fig. 1.1). The three-dimensional folding funnel, however, can be more simply rendered as a two-dimensional reaction coordinate where the enthalpic and entropic terms are combined to give an activation free energy, or ΔG^\ddagger . The uphill struggle, therefore, opposes the large number of possible conformational states, and the reaction proceeds downhill once the stabilizing hydrophobic effect contribution exceeds the entropic penalty.

1.4.3 Hydrogen bonds

Considerable debate (BenTal *et al.*, 1997; Dill, 1990; Kentsis & Sosnick, 1998; Khare *et al.*, 1999; Krantz *et al.*, 2000; Pace *et al.*, 1996; Shi *et al.*, 2002; Yang & Honig,

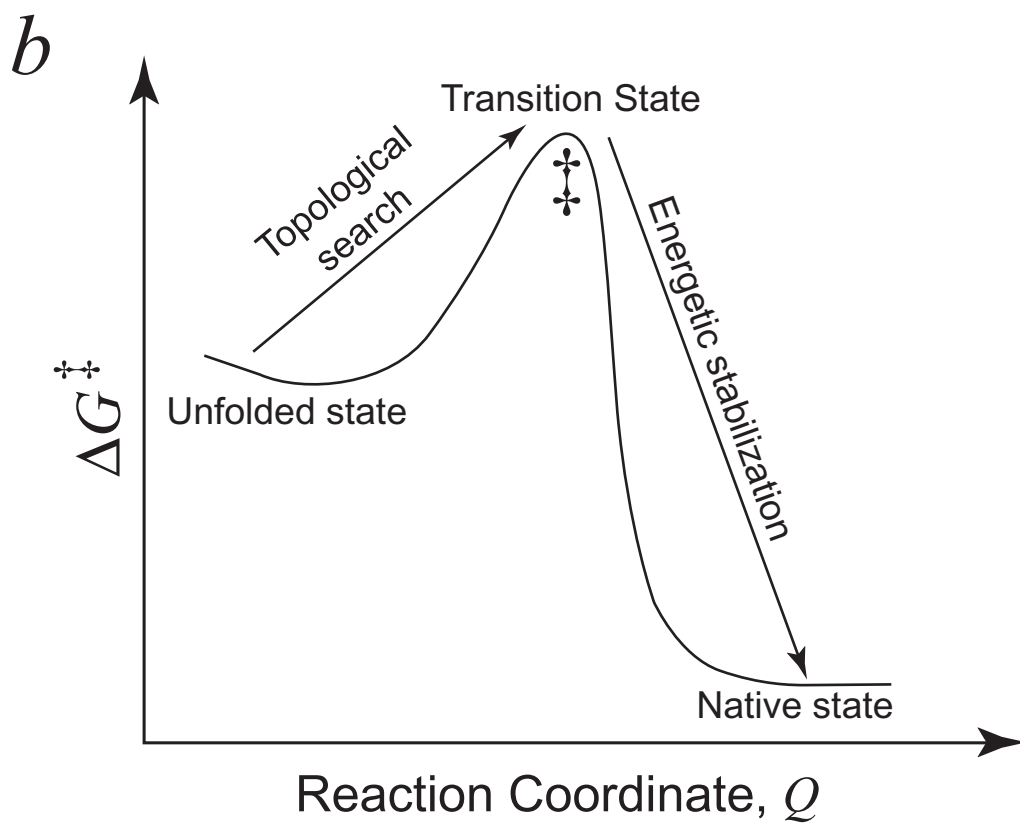
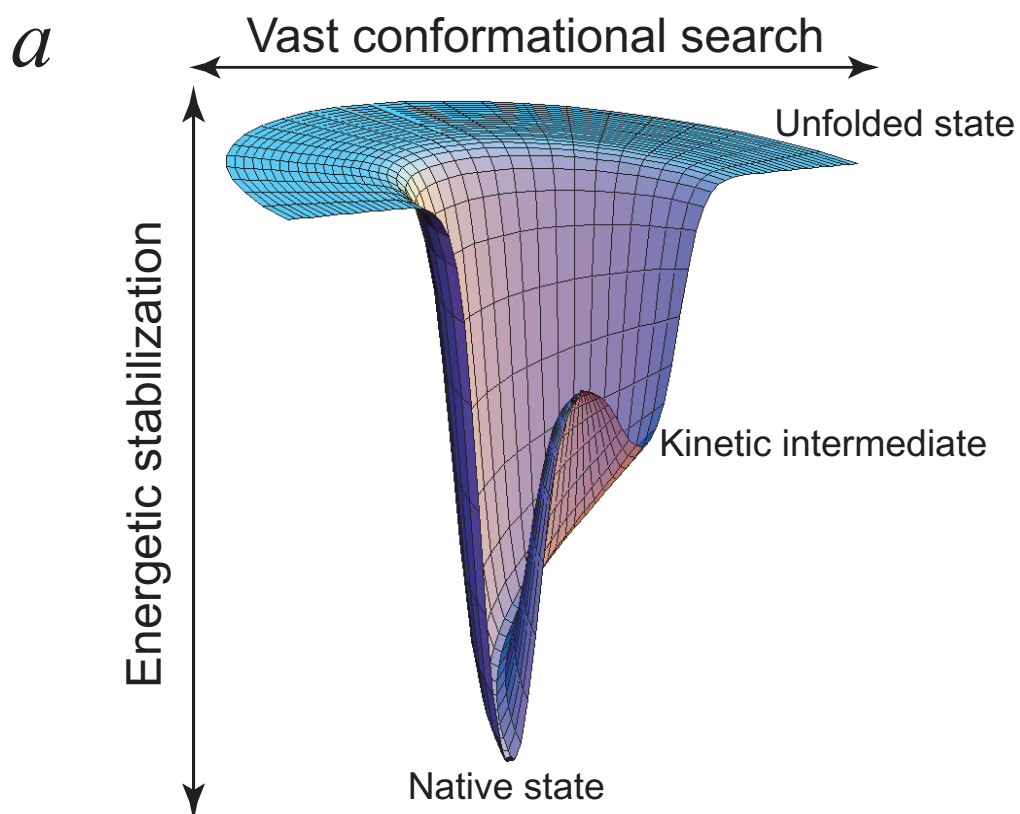
1995) surrounds the energetic contribution of the hydrogen bond (H-bond). When closely examining the details of protein H-bond formation, the unfolded amides and carbonyls are solvated with water and thus require backbone desolvation to form native H-bonds. It is expected that the free energy gained from forming a native protein amide-carbonyl H-bond is lost to the desolvation of these moieties in the unfolded state (Dill, 1990; Yang & Honig, 1995). D/H isotope effect measurements (Krantz *et al.*, 2000; Krantz *et al.*, 2002) as well as fractionation factor measurements (Bowers & Klevit, 1996; Bowers & Klevit, 2000; Khare *et al.*, 1999; Loh & Markley, 1993; Loh & Markley, 1994) confirm that the solvated unfolded state for an H-bond is nearly equal in energy to the folded state. This topic will be rigorously discussed in Chapter 3.

An alternative manner of explaining the role of the protein H-bond energetics is that while H-bonds do not contribute to the net stability of the protein, the H-bond is not readily broken once formed deep inside a protein core, because of the large penalty associated with burying an unsatisfied polar functional group within a non-polar environment. Thus H-bonds guide the formation of protein structure and encode structural specificity. Because once formed, the structure cannot readily slip into an alternate conformation, e.g. a β -sheet may not simply shift out of register without breaking H-bonds in an unfavorable, non-polar environment.

1.4.4 Electrostatics

Electrostatic energies are often the result of the pair-wise application of Coulomb's Law to charged functional groups located on a protein's surface (Horovitz *et al.*, 1990; Serrano *et al.*, 1990; Warshel & Russell, 1984). A dielectric term is introduced to account for the relative differences in polarizability of aqueous solvent and protein

Figure 1.1 Typical depictions of protein folding reaction coordinates. **a**, The three-dimensional funnel depiction of a protein folding reaction coordinate represents the broad number of conformations in the unfolded state as the widest upper portion of the funnel. As the protein ensemble is energetically stabilized, the number of possible conformers is reduced and the funnel narrows until reaching the singular native state. Kinetic intermediates may be populated on the landscape surface, but not all intermediates are required for folding. **b**, A simplified two-dimensional reaction coordinate that is defined by the activation free energy of the folding protein, however, is often adequate in describing the folding pathway of two-state folding proteins. Here, the topological search describes the uphill struggle against conformational entropy, where the transition state is defined as the species (or ensemble of species) that contains the minimal subset of contacts such that any additional interactions stabilize the ensemble. Thus, the downhill portion represents the energetic stabilization as the folding protein's nascent nucleus is buttressed by additional surface burial while achieving the native state.



cores (ϵ of 80 and 4-8, respectively). Oppositely charged functional groups on the surface stabilize a protein at the expense of the entropic penalty associated with immobilizing side chains. Furthermore, like charged groups repel one another. Therefore, it has been observed that thermophilic enzymes have adapted thermostability from electrostatic energetics by *minimizing* the number of electrostatic repulsions rather than *maximizing* the number of electrostatic attractions, since the latter attractions are at the expense of side chain conformational entropy (Loladze *et al.*, 1999; Perl & Schmid, 2001). Electrostatic studies on ubiquitin revealed similar findings (Ibarra-Molero *et al.*, 1999; Loladze *et al.*, 1999; Loladze & Makhatadze, 2002).

1.4.5 Van der Waals

Van der Waals energies are derived from classic Leonard-Jones potentials where the optimal distance between atoms is approximately the sum of the van der Waal radii. The free energy stabilization afforded by the depth of the well contributes modestly to the protein's overall stabilization. However, the unfolded state itself contains a large number of van der Waal interactions with water solvent. Therefore, like the H-bond energetic contribution, the net van der Waal contribution may provide little added stability due to a cancellation of terms.

1.5 Surface burial

Unlike the various energies described in the previous section, surface burial, as with the relaxation time for folding, is a readily measured equilibrium and kinetic quantity. For ubiquitin, about 6100 Å² are buried in the folding process. Tanford and

others revealed that equilibrium denaturant dependence of protein folding correlates linearly with denaturant concentration (Aune & Tanford, 1969; Tanford, 1968; Tanford, 1970). This constant dependence is called an m -value. Jackson and Fersht showed that for small two-state folding proteins, like CI2, the folding and unfolding activation energies also varied linearly with denaturant, and these m -values are often denoted by m_f and m_u , respectively (Jackson & Fersht, 1991a). Pace and co-workers noted that the proportionality between guanidinium chloride and urea denaturants are $0.22 \text{ kcal mol}^{-1} \text{ M}^{-1} \text{ \AA}^{-2}$ and $0.11 \text{ kcal mol}^{-1} \text{ M}^{-1} \text{ \AA}^{-2}$, respectively (Myers *et al.*, 1995). Thus, a direct relationship was established between the denaturant dependence m -value and protein surface burial during folding (App. 3).

Equilibrium heat capacity measurements of protein folding stability have provided even greater detail of the surface burial component of the folding reaction. These studies were conducted on a number of proteins and indicate that protein surface can be divided into polar and non-polar surface where the latter contributes two parts favorably to ΔC_p and the former destabilizes it by one part (Loladze *et al.*, 2001; Makhatadze, 1998; Makhatadze & Privalov, 1990; Privalov & Makhatadze, 1990).

1.6 Modeling protein folding

How the conformational freedom of a polypeptide is restricted during folding is a topic of considerable debate, stemming from both theoretical predictions and conclusions from experimental studies. Several models elaborate on the events that define the steps leading to the rate-limiting species.

1.6.1 Diffusion-collision and hydrophobic collapse

The *diffusion-collision model* (Bashford *et al.*, 1988; Burton *et al.*, 1997; Burton *et al.*, 1998; Karplus & Weaver, 1994; Myers & Oas, 1999; Myers & Oas, 2001), or *framework model*, (Kim & Baldwin, 1990) state that local secondary structures form early in the folding pathway and these pre-formed structures diffuse and collide forming late tertiary contacts, or adhesions (Fig. 1.2). Adherents to this model believe that early secondary structure formation is crucial to reducing the conformation search. However, many proteins, such as ubiquitin prefer to form the improper secondary structures in isolation; and likewise, secondary structure prediction algorithms, such as AGADIR, fail to correctly predict helical structure in native ubiquitin (Munoz & Serrano, 1994; Munoz & Serrano, 1997). Moreover, the fact that proteins fold faster at higher temperatures is inconsistent with a rapid pre-equilibrium in which secondary structure is formed. Diametrically opposed to these models is the *hydrophobic collapse model* (Ptitsyn, 1995) which contends that hydrophobic side chains rapidly coalesce forming a nascent core of distant tertiary contacts (Fig. 1.2). Upon this hydrophobic core nucleus, local secondary structures propagate. These models differ exactly in the temporal order of secondary and tertiary structure formation along the folding pathway.

These extremely polarized theories about the nature of the folding pathway are probably overstated, because not all proteins will fold according to the same set of rules, and most likely a pathway specific to a protein will combine aspects of local secondary structure formation with critical tertiary contacts to define the transition state. Chapter 3 partially addresses the merits of these models in the context of hydrogen bond formation and surface area burial.

1.6.2 Search nucleation

While the concept of nucleation was initially posited from early experimental studies (Ristow & Wetlaufer, 1973; Wetlaufer, 1973), it fell out of favor as the likely mechanism of protein folding, since accumulated intermediates were observed (Kim & Baldwin, 1982). Albeit, such intermediates populate due to proline isomerization, disulfide bond shuffling, or histidine-heme mis-ligation. Such intermediates accumulate when misfolding by such processes requires a slow reorganization/repair step (Sosnick *et al.*, 1994). Fersht and co-workers revealed that intermediate accumulation is non-essential when CI2 was shown to fold without populated intermediates (Jackson & Fersht, 1991a). Several years later, Fersht re-introduced the concept of nucleation to explain the two-state folding pathway for CI2 (Fersht, 1995).

Contemporary to Fersht's nucleation model, the concept was also proposed as *search nucleation*, (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995) which posits that neither pre-formed helix nor excess hydrophobic collapse dominates the early steps of protein folding (Fig. 1.2, *bottom*). Instead what limits the process of protein folding is a *search* for the minimal number of contacts which define the native topology. This search is accomplished by a trial-and-error process wherein the protein forms and releases from transient topology pinning hydrophobic interactions, which contain crucial structure determining H-bonds (Krantz *et al.*, 2000; Krantz *et al.*, 2002). When unfolded polypeptide is greatly divergent from the native topology, it rapidly melts away due to its weak stability. If, however, the topology is correctly formed, then this nucleated structure has achieved the rate-limiting step and proceeds energetically downhill to the native state.

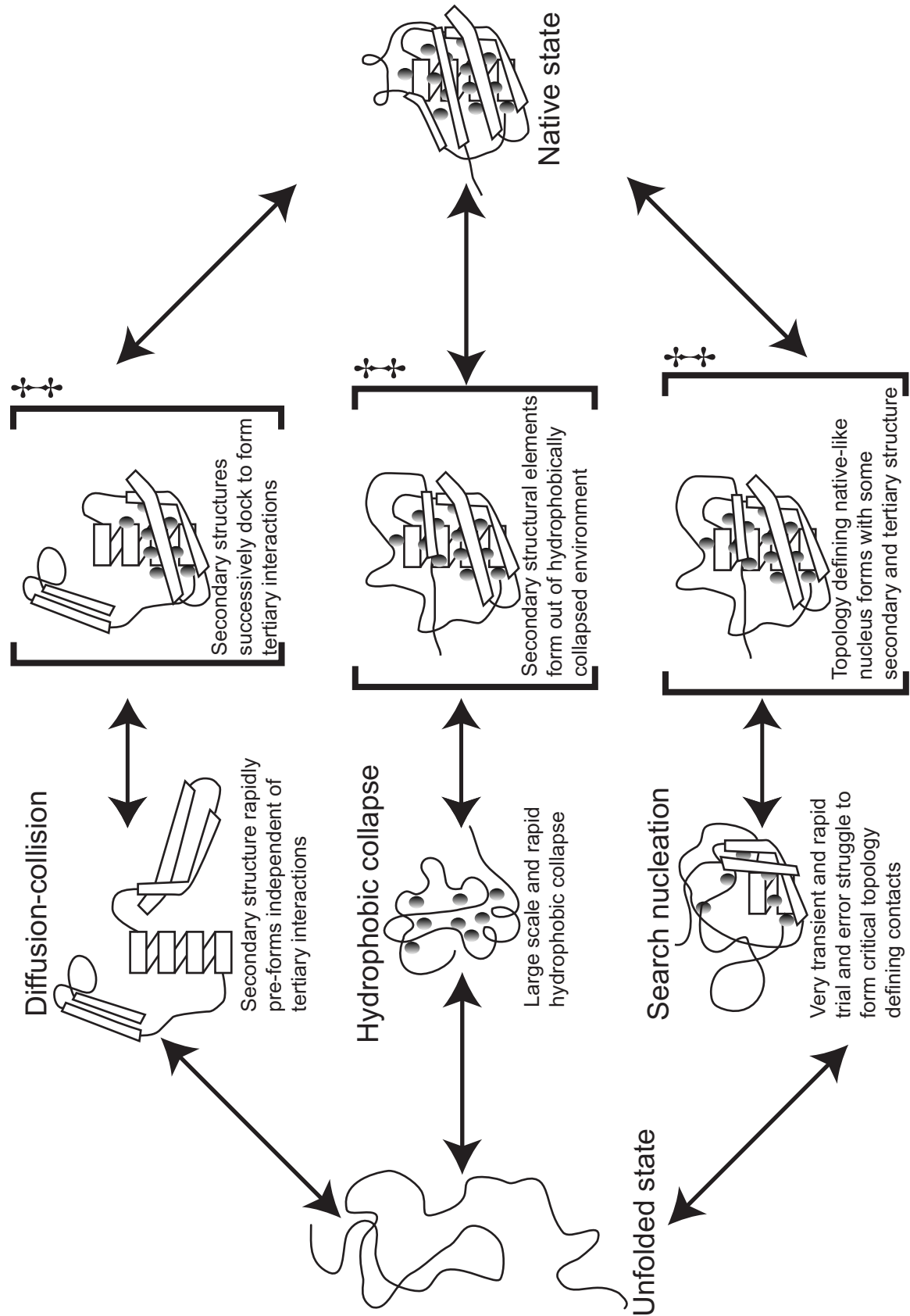
Since the topological complexity of a particular protein is directly related to the protein's folding speed (Plaxco *et al.*, 1998), the rate in which protein of a particular topology can sample topologies by this trial-and-error process should be constant. Slower folding proteins with more tortuous and complex topologies may have to search through many more possible topologies to find the native-like topology than proteins with less complex folds. Put another way, the search may take less time for less complex proteins, because the topology defining interactions are nearer in primary sequence space.

Experimental evidence for the search nucleation model is apparent from studies that mutate residues in turns (Grantcharova *et al.*, 2000; Grantcharova *et al.*, 1998; Gu *et al.*, 1997; Martinez *et al.*, 1998; McCallister *et al.*, 2000; Munoz *et al.*, 1997; Nauli *et al.*, 2001). These turn mutants are more critical to the rate-limiting step than those residues often found within contiguous secondary structure; and therefore, a turn which defines the topology is more rate-limiting than simple secondary structure formation of helices or sheet strands. When a coiled coil system was studied, it was shown that the terminus with the greatest helical propensity, in fact, folded late in the pathway. The structure which was rate-limiting, while possessing weaker helical propensity, was nearer the cross-link crucial to defining the coiled coils overall topology (Krantz & Sosnick, 2001; Moran *et al.*, 1999).

1.6.3 *Are there intermediates?*

The search nucleation folding model implies that stable folding intermediates do not populate along the folding pathway. Here stable folding intermediates refer to species that are more stable than the unfolded state but exist prior to the rate-limiting barrier (Krantz & Sosnick, 2000). Only such intermediates are spectroscopically

Figure 1.2 Putative models for protein folding. Three potential folding models are depicted where the earliest steps immediately after the unfolded state largely distinguish each model. The rate-limiting structure in brackets is designated by the symbol (\ddagger). The Diffusion-collision Model (*top*) pre-supposes that the secondary structural elements rapidly form independent of distant tertiary interactions. These secondary elements productively associate in the rate-limiting structure. The diffusion limit describes the maximal rate at which these structures can interact. The Hydrophobic Collapse Model (*middle*) diametrically opposes the diffusion-collision model in that the tertiary interactions made by rapid hydrophobic collapse dominate the early events. Thus secondary structural elements emerge out of the collapsed species. In the Search Nucleation Model (*bottom*), the unfolded protein rapidly samples a multitude of geometries with both secondary and tertiary structure. Failed nucleation events rapidly melt back to the unfolded state. Those nucleated structures that best define the gross topology of the native protein achieve the rate-limiting step.



observable if they accumulate during refolding. The Initial Barrier Hypothesis, however, may be combined with the search nucleation model (see Ch. 2). As nearly fifty small globular proteins (~110 residues) fold in a two-state manner (Jackson, 1998), it is believed that the search for the correct topology is the major early limiting barrier for folding (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995). The only significantly populated intermediates for cytochrome *c*, for example, occur after the early search nucleation barrier (Bai *et al.*, 1995b).

Therefore, despite the tremendous complexity of proteins, the many thousands of atoms and interactions that comprise the singular native structure of a small globular protein fold from a disordered unfolded state in a seamless two-state transition where only the native and unfolded states are sigmoidally populated. This seemingly paradoxical experimental realization perplexes theorists and experimentalists alike. However, because an overwhelming majority of many small globular proteins are two-state folders, it is readily conceded that a high degree of cooperativity moderates the folding process, precluding the transient population of partially folded forms prior to the major rate-limiting barrier (excluding *cis/trans*-proline isomerization, disulfide reshuffling, misfolding, and protein aggregation). This topic will be dealt with extensively in Ch. 2, addressing the apparent exceptions, which fold through putative on-pathway intermediates.

Generally speaking, proteins larger than ~110 amino acids can possess more than one domain. In these cases, multiple folding intermediates may populate readily, because either domain may fold independent of the other. Thus the Initial Barrier Hypothesis may still apply at the individual domain level in such systems as hen egg

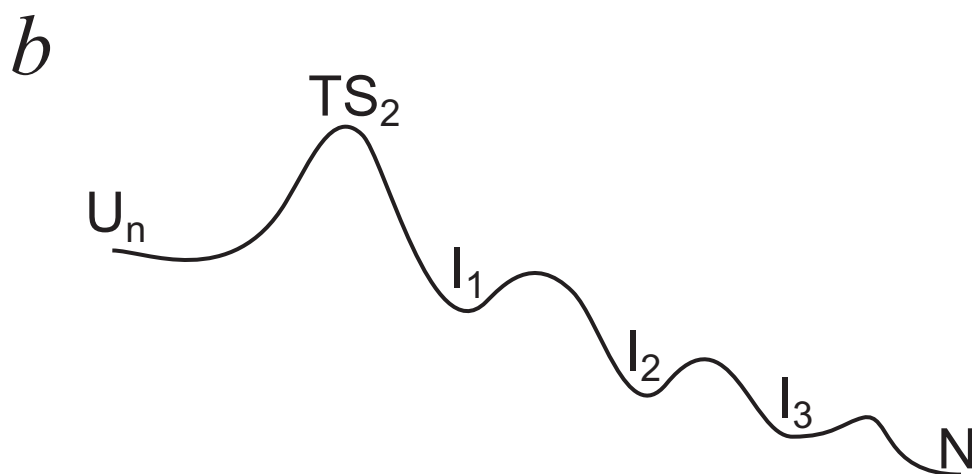
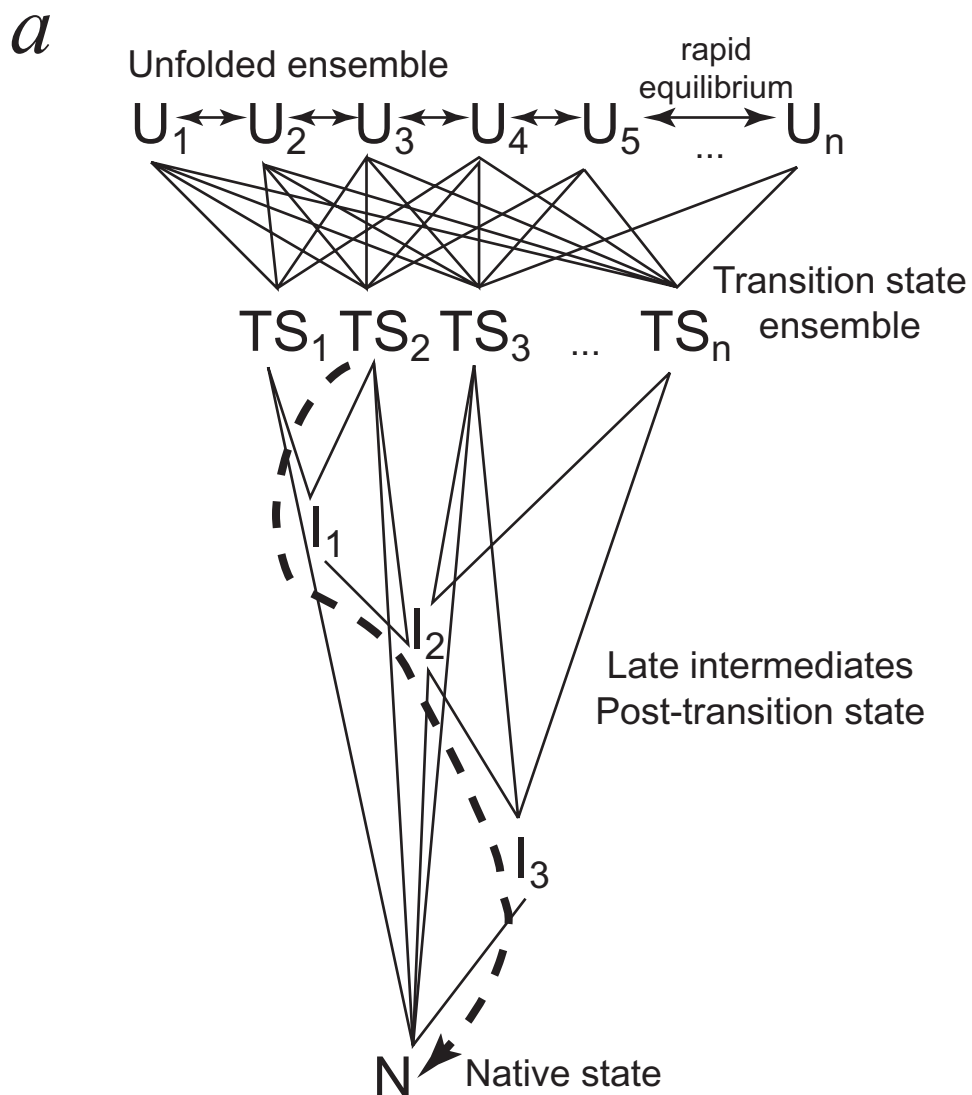
white lysozyme (Radford *et al.*, 1992; Wildegger & Kiefhaber, 1997) and RNase H (Chamberlain *et al.*, 1996; Parker & Marqusee, 1999; Raschke *et al.*, 1999) despite the fact that they appear to populate kinetic intermediates.

1.6.4 Pathways or landscapes?

Protein folding is often considered a determinate process, whereby specific intermediates populate along a singular pathway. Most folding experiments (Fersht *et al.*, 1994; Grantcharova *et al.*, 2000; Martinez *et al.*, 1998), but not all (Burton *et al.*, 1997; Moran *et al.*, 1999), are interpreted in the context of a homogeneous transition state ensemble, while theoretical work, on the other hand, has led to a funnel picture in which folding occurs via structurally distinct, heterogeneous routes (Dill & Chan, 1997; Shakhnovich, 1998; Socci *et al.*, 1998; Thirumalai & Klimov, 1998). This controversy, “the classical versus new view debate”, has become central to the protein folding pathway discourse (Baldwin, 1995; Dill & Chan, 1997).

The contention surrounding this issue is whether a protein traverses the same barrier each time it folds to the native state. Moreover, does the protein populate the same sequence of structures along a given pathway each time it folds? Or does a protein gradually build up structure from a variety of nucleation sites where the connectivity of the various intermediate structures are better described by a multi-dimensional landscape rather than a linear determinant process? This topic will be dealt with in greater detail in Chapter 4 (see Fig. 1.3).

Figure 1.3 Branched landscapes versus linear folding pathways. Simple diagrams depicting the intrinsic differences between the landscape model and the linear pathway model. **a**, In the landscape model, the unfolded ensemble (U) rapidly equilibrates, folding through multiple distinct rate-limiting species (TS). Even for a spectroscopic two-state folding protein, multiple kinetic routes may exist where the most often traversed routes are the most stable. Conversely, the linear pathway model (**b**) describes the folding pathway as a dominant route which is so highly preferred that the peripheral routes, while possible, are negligible. Therefore, a highly complex folding landscape may exist in theory but in practice only a single route is probable (..... in **a**).



2

The Initial Barrier Hypothesis In Protein Folding

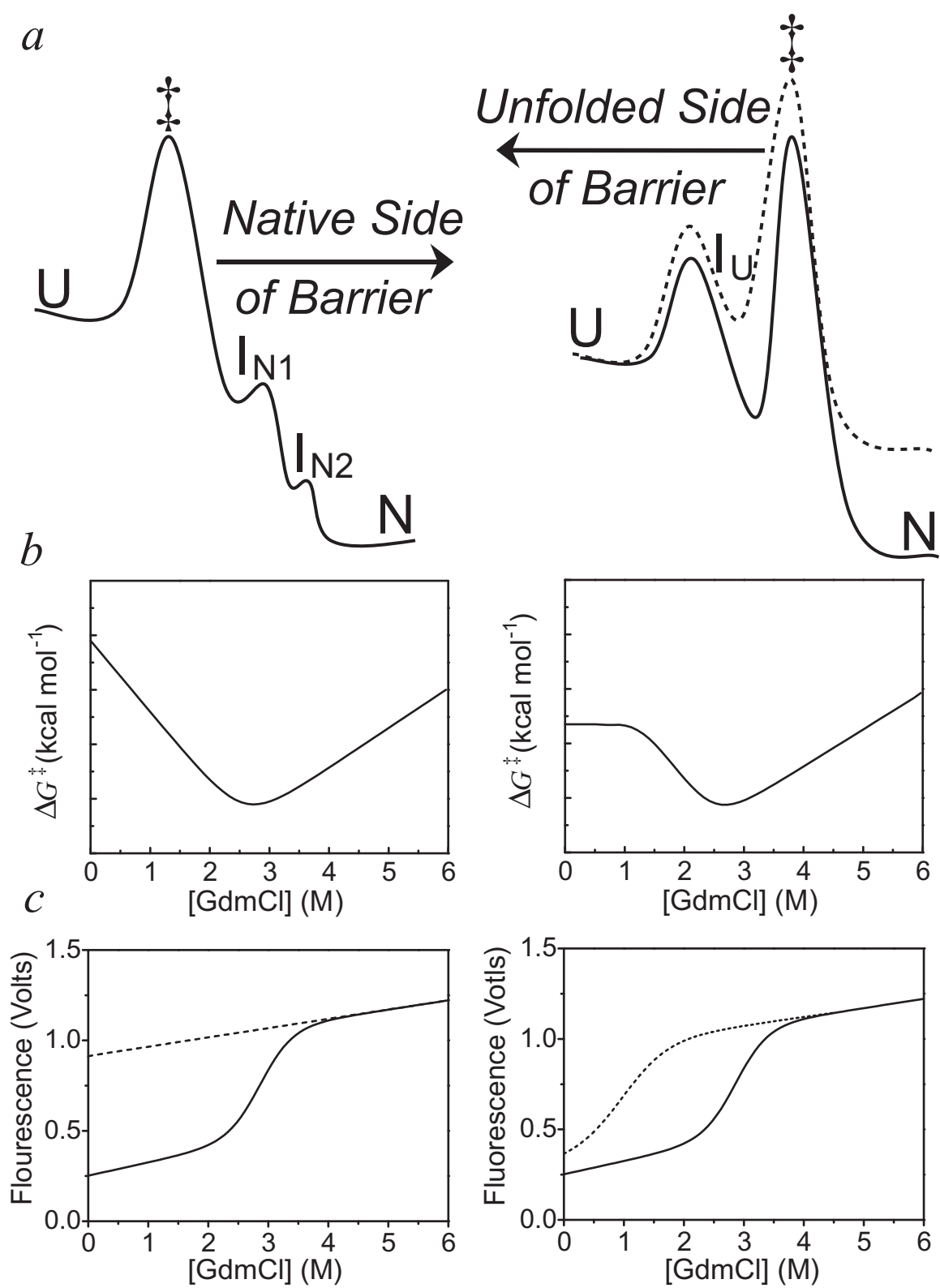
2.1 Abstract

The folding strategy of proteins has been suggested to involve an initial, large scale, rate-limiting nucleation event followed by late smaller barriers; extensively organized intermediates accumulate only when misfolding inserts a large error-repair barrier after the initial committed step. A prediction from this view is that small proteins in particular should tend strongly to fold in a two-state manner, without such intermediates accumulating. The generality of this hypothesis is discussed.

2.2 Introduction

In spite of the high dimensionality of the reaction surface, many small globular proteins kinetically fold in a two-state manner: $U \leftrightarrow N$ (Jackson, 1998). This observation is powerful, because it places strong constraints on any relevant folding model. For example, two-state behavior mandates that any intermediate more stable than the denatured state must exist on the native-side of the rate-limiting barrier under refolding conditions (Fig. 2.1). Based upon this and other observations, it is proposed that the folding of small globular proteins (less than 100 amino acids) is a nucleation process (Abkevich *et al.*, 1994; Fersht, 1995; Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995; Thirumalai & Guo, 1995) with a limiting step that is relatively early in the folding process (Abkevich *et al.*, 1994; Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995; Thirumalai & Guo, 1995).

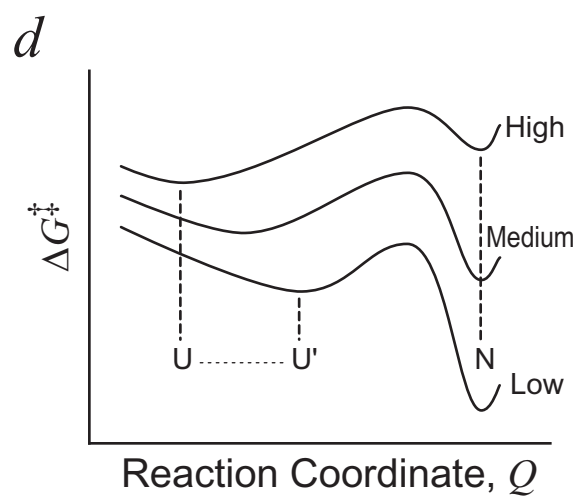
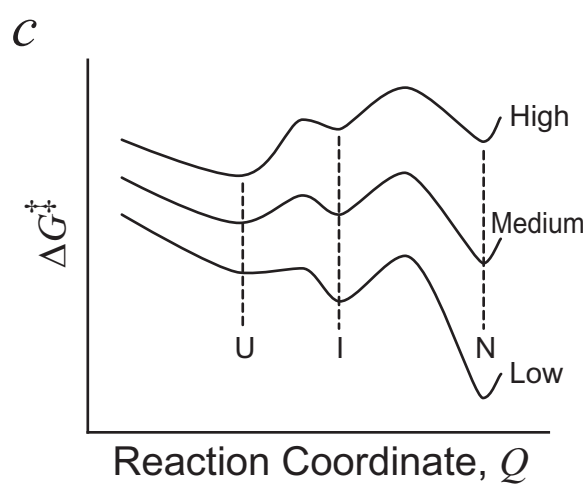
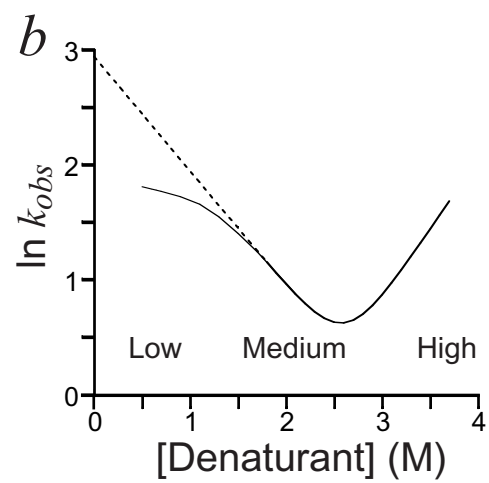
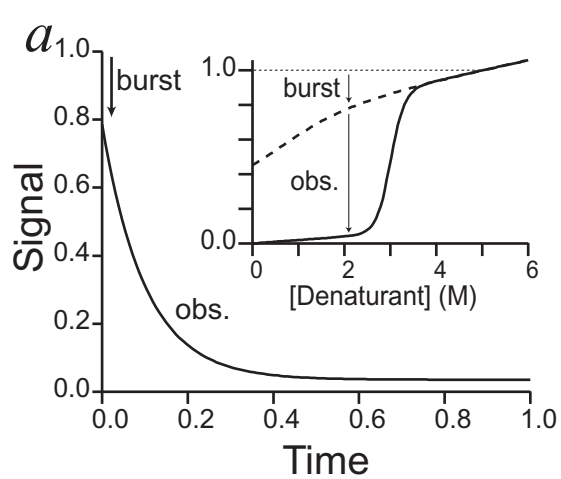
Figure 2.1 Two-state versus three-state folding. Folding behavior for model spectroscopic two-state (*left*) and three-state (*right*) proteins. **(a)** Free energy diagrams. An intermediate, I_U , occurring prior to the rate limiting step (\ddagger) is observed in refolding experiments when it is stable relative to the unfolded state. **(b)** Denaturant dependence of standard free energy of activation. Roll-over can be described as the three-state region of the chevron (— in *a on right*) while the remainder of the chevron is predominantly two-state when denaturant destabilizes I_U relative to U (\cdots in *a on right*). The three-state plot shown assumes that the amount of surface buried in I_U is the same as in the transition state. **(c)** Initial and final values of the observed signal. The final amplitude (—) represents the equilibrium values. In the three-state scenario at low denaturant concentrations, there is a significant loss of signal for the initial value (\cdots) with a sharp sigmoidal dependence indicative of a cooperative U -to- I_U melting transition.



In earlier work, evidence indicated that the intrinsic rate-limiting barrier in protein folding occurs as the initial on-pathway step (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995). Later barriers can slow folding and lead to the accumulation of folding intermediates. But these barriers are adventitious, due to the repair of misfolding errors rather than to difficulties inherent to the folding process (Sosnick *et al.*, 1994). The initial barrier in cytochrome *c* (Cyt *c*) was characterized. The initial barrier represents, in terms of free energy, a large scale, uphill conformational search for some extended nucleus that is native-like in chain winding topology (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995). The native-like topology reached in this initial step can, for the first time, guide and stabilize the formation of native-like intermediates in a downhill manner. Subsequent stepwise formation of native structure, guided and stabilized in turn by prior structure (Rumbley *et al.*, 2001) requires fast, small scale conformational searches and tends not to be rate-limiting. Thus, in the absence of late adventitious barriers, intermediates with long-range structure do not accumulate; folding appears to be two-state and limited by the initial nucleation step.

This “Initial Barrier Hypothesis” leads to some interesting predictions. One expectation concerns the probability for proteins to fold in either a two-state or a three-state manner (Fig 2.1). Stable intermediates with more than just some local, residual structure (e.g. turn or helix) cannot form before the obligatory initial nucleation step. Later intermediates can be stable and obligatory, but tend not to accumulate, because they form after the rate-limiting step. Misfolding barriers suggested to cause intermediate accumulation are especially unlikely for small proteins: the extended native-like topology, which is pinned into place by the large scale nucleation process, will represent such a large fraction of the chain that folding errors should be improbable.

Figure 2.2 The burst-phase and the chevron roll-over in protein folding. (a) Illustrative kinetic trace for folding after denaturant dilution, showing a fast burst-phase and a slower observable phase. The dependence on denaturant of both phase amplitudes is illustrated in the inset. *(b)* Illustrative kinetic chevron showing the denaturant dependence of folding and unfolding rates, with a roll-over in the rate of folding to N at low denaturant (reduced m -value; rates less than expected). *(c,d)* Alternative folding free energy profiles that might explain initial burst-phase and roll-over behavior. *(c)* A discrete intermediate is pictured to form in a fast barrier-crossing event (Eq. 2.1), producing burst-phase signals (small barrier) and folding rates slower than expected (roll-over) at low denaturant. *(d)* The unfolded molecules when placed into native conditions are pictured to relax initially to a more contracted but still unfolded ensemble (Eq. 2.2), producing burst-phase signals and roll-over.



Initially it is quite surprising to believe that any protein could fold in a two-state manner without populating intermediate forms, yet nearly forty small proteins are now known to do so (Jackson, 1998), consistent with our expectation. However, counter to our expectation, other small proteins have been interpreted to populate very early “burst-phase” intermediates, apparently before the early barrier, and also late intermediates due to large obligatory on-pathway barriers. In this chapter, data are presented that examine many of these apparent exceptions and provide an alternative explanation for sub-millisecond burst-phase signals, revealing that nearly all small proteins fold in a manner which is consistent with the Initial Barrier Hypothesis.

A growing number of proteins appear to populate stable on-pathway intermediates (Eq. 2.1) on a fast, sub-millisecond time scale, before the rate-limiting step in folding (Eaton *et al.*, 1997; Roder & Colon, 1997). Here, the term “intermediate” is used to describe species with some secondary and tertiary structure. Such species are to be distinguished from conformations containing some regions of stable local turn or helical structure, which may best be described as residual structure, which should be considered part of the denatured ensemble.

Intermediates which accumulate can be directly detected using ultra-fast techniques, or in stopped-flow experiments as a “burst-phase” missing amplitude at low denaturant (Fig. 2.2*a*). Proteins with accumulated intermediates also tend to exhibit a roll-over decrease in folding rate at low denaturant (Fig. 2.2*b*). The relationship of the burst-phase and roll-over behavior to intermediate accumulation has been rationalized in terms of Fig. 2.2*c* and Eq. 2.1 (Khorasanizadeh *et al.*, 1993; Matouschek *et al.*, 1990). In this view a rapidly formed on-pathway intermediate becomes stable at low denaturant

concentration and accumulates in a fast phase (burst-phase), before the rate-limiting barrier, causing folding to become slower than expected (roll-over). Alternatively, as in Fig. 2.2*d* and Eq. 2.2, the burst-phase and roll-over may simply reflect some solvent-dependent relaxation of the U state ensemble, such as chain contraction (Sosnick *et al.*, 1996b) or interchain aggregation (Silow & Oliveberg, 1997), promoted by random side chain association when the unfolded polypeptide is mixed from high denaturant (good solvent) into lower denaturant (poor solvent).



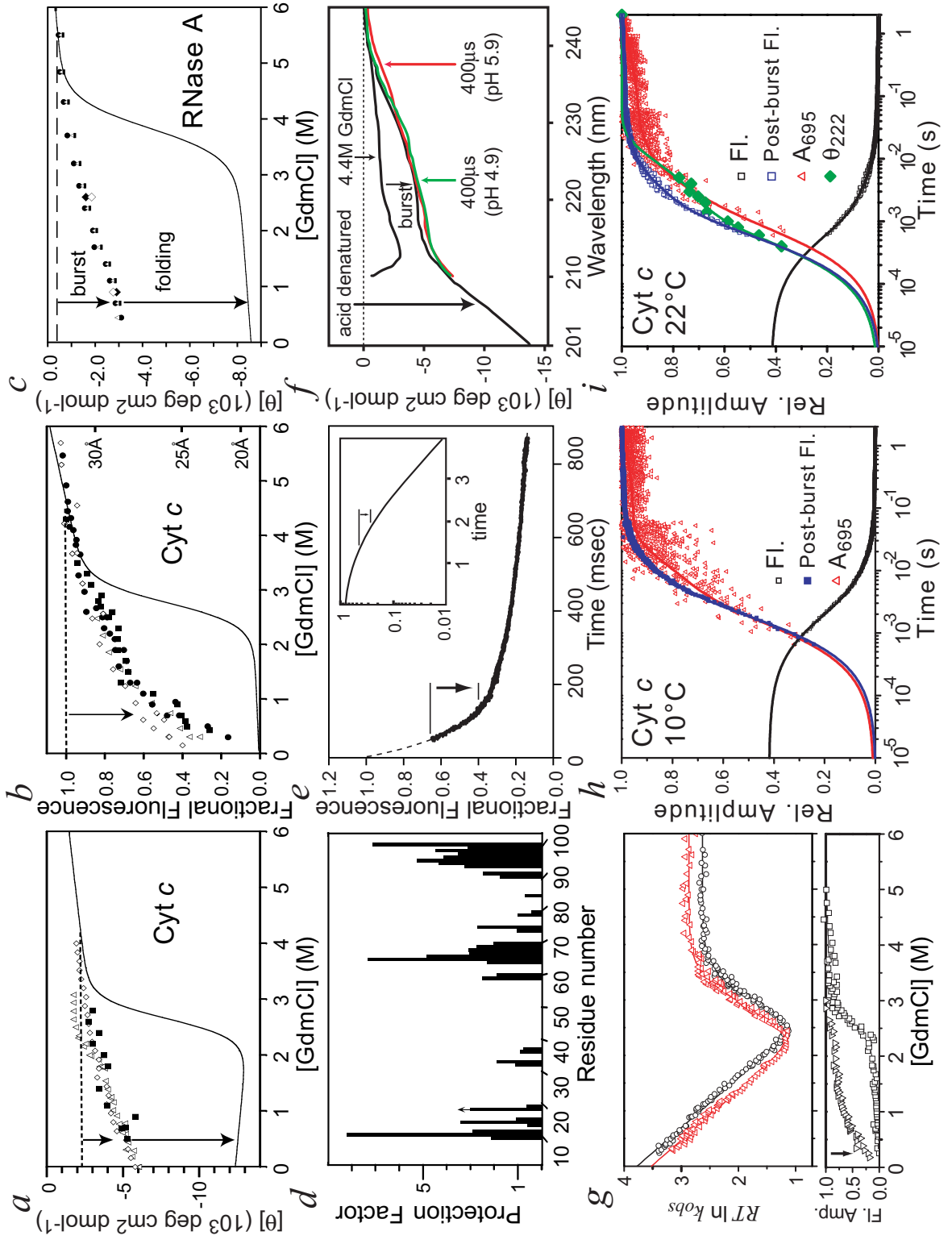
Eqs. 2.1 and 2.2 are formally identical and difficult to distinguish experimentally. In both cases, the altered state in kinetic folding experiments will appear “obligatory” and “on-pathway” while producing burst-phase and, in certain cases, roll-over behavior. Subsequently, folding will be mono-exponential (Parker & Marqusee, 1999). However, these cases are importantly different in their structural and thermodynamic implications for protein folding processes (compare Figs. 2.2*c,d*).

Evidence favoring solvent-dependent U state readjustment has been found for Cyt *c* (Chan *et al.*, 1997; Sosnick *et al.*, 1997) and ribonuclease A (RNase A) (Qi *et al.*, 1998). In the Cyt *c* example (Fig. 2.3*a,b*), two large fragments (1-65 and 1-80) which lack the most stable C-terminal helix segment (residues 86-103) were used as non-folding models to calibrate the circular dichroism (CD) and fluorescence (Fl) of the U state under native, low denaturant conditions (Sosnick *et al.*, 1997). For RNase A (Fig. 2.3*c*), the non-folding disulfide-broken molecule served as an unfolded state model (Qi *et al.*, 1998). Figs. 2.3*a-c* show the equilibrium denaturant dependence of these molecules. The

intact proteins melt and refold cooperatively. The non-folding analogs readjust their U state ensemble in a continuous, non-cooperative way from more extended chains at high denaturant (good solvent) to more contracted chains at low denaturant (poor solvent). Figs. 2.3*a-c* also show kinetic results when these molecules are diluted rapidly from denaturing conditions. The CD and FI parameters of the U state models change from their equilibrium high denaturant values to their equilibrium low denaturant values in a sub-millisecond burst-phase. The intact proteins quantitatively duplicate the burst-phase jump seen for the non-folding analogs in the same non-cooperative way over the whole range of denaturant concentration; however, intact proteins fold to N at a slower time scale. This rather striking correspondence suggests that, like the U state models, the solvent change simply readjusts the still unfolded proteins to their new equilibrium unfolded condition rather than to a structured intermediate, producing burst-phase signals.

Additional evidence favoring solvent-dependent U state relaxation has been found for barstar (Agashe *et al.*, 1995; Bhuyan & Udgaonkar, 1999), RNase H (Parker & Marqusee, 2000), and lysozyme (Chen *et al.*, 1998) and by theoretical analysis (Gutin *et al.*, 1995). In conflict, various studies with other proteins have been interpreted in favor of distinct structured burst-phase intermediates, before the nucleation barrier, as in Eq. 2.1 and Fig. 2.2*c*. This chapter considers the latter and some pertinent new results. I then consider later intermediates that accumulate after the initial barrier has been passed.

Figure 2.3 Cyt *c* folding. Equilibrium and kinetic denaturant dependence of CD (**a**) and Fl (**b**) for intact Cyt *c* and two non-folding analogs (fragments 1-65 and 1-80). Solid lines show equilibrium melting curves for the intact protein. Open symbols define analogous melting curves for the non-folding analogs, obtained both at static equilibrium and after their kinetic burst-phase upon denaturant dilution. Closed symbols show the burst-phase for the intact protein (upper arrows), and lower arrows indicate its slower folding to N (from Sosnick *et al.*, 1997). **c**, Analogous results for RNase A using the protein with its four disulfides broken for a non-folding reference (from Qi *et al.*, 1998). **d**, HX protection factors for equine Cyt *c*, measured by pH competition during 2 msec of folding at elevated pH in 0.4 M Na₂SO₄, 0.3 M GdmCl, 10°C (from Sauder & Roder, 1998). **e**, Fast Fl quenching data for Cyt *c* after a jump from pH 2, low salt (unfolded) to pH 4.5 (native conditions). Data from (Shastry & Roder, 1998). Data fitting includes a putative mono-exponential burst-phase with time constant of 50 μ sec (*arrow*). The inset is a simulation of the Fl quenching signal computed for a continuous, non-barrier-crossing chain collapse (*as in d*) starting with Trp59 far from the heme (Hagen & Eaton, 2000). Horizontal lines mark the measured range of the ~ 50 μ sec burst-phase. **f**, CD spectra for early Cyt *c* folding measured in fast continuous flow mode (from Akiyama *et al.*, 2000). The initial CD spectrum in high denaturant is shown. Upon denaturant dilution, Cyt *c* relaxes within 400 μ sec to a condition with essentially the same CD spectrum as the acid denatured state (shown), the thermally denatured state, and the non-folding fragments (**a,b**), which duplicate the spectrum identified by Akiyama *et al.* as Intermediate 1 (<10% helix). When initial conditions used acid unfolding, CD starts from the acid denatured spectrum and its measured time dependence (after 400 μ sec) extrapolates back to that spectrum, i.e. no pre-400 μ sec burst-phase is seen (see Fig. 6 of Akiyama *et al.*, 2000). **g**, Chevron plots for mutated Cyt *c*. The replacement of the two peripheral histidines of Cyt *c* (H26N, H33N) avoids initial aggregation and shows no chevron roll-over in folding, even though the contraction-dependent Fl burst-phase is still present (lower panel, arrow). The two chevrons are for ^{NH}Cyt *c*^{PWT} (○) and NDCyt *c*^{PWT} (△) in H₂O, where the latter are deuterated at exchangeable amides (see Ch. 3). The H/D isotope effect shows that the majority of the H-bonds (3 helices) form in the rate-limiting step leading to N and not in the prior burst-phase (Krantz *et al.*, 2000). Two-state Cyt *c* folding, with some misfolding-dependent heterogeneity (small slow phase). Folding kinetics are measured by Fl and A₆₉₅ absorbance at 10°C (**h**) and 22°C (**i**) after a pH jump (pH 2 unfolded jumped to pH 4.5 for folding). The observed Fl amplitude (descending curves, burst-phase $\sim 60\%$) is normalized to total Fl. For direct comparison with A₆₉₅ (*red*), the kinetically observable Fl (*ascending curves, blue*) is normalized to the total kinetically observable Fl decay, omitting the Fl burst-phase. Green points (**i**) are from previously acquired CD data (Akiyama *et al.*, 2000). Post-burst-phase folding measured by the initial collapse probe (Fl) and the final N-state probe (A₆₉₅) are identical (**h**), or nearly so (**i**), showing that the major folding phase, which matches k_{middle} of Akiyama *et al.*, represents 2-state folding to N and not to some intermediate. A minor slow phase detected selectively by the native state probe (A₆₉₅) tracks the folding of a separate small population (Akiyama *et al.*, 2000) that encounters an adventitious misfolding barrier. Conditions are identical to previous work (Akiyama *et al.*, 2000).



2.3 Early intermediates

2.3.1 Cytochrome *c*

When Cyt *c* folding is initiated by denaturant dilution or a pH jump, Fl is quenched in a sub-millisecond burst-phase due to Förster transfer as the lone tryptophan residue (Trp59) is brought close to the heme covalently bound to Cys14 and Cys17. Two types of evidence support the conclusion that a distinct folding intermediate is formed in the Fl burst-phase. Sauder and Roder found widespread hydrogen exchange (HX) protection in Cyt *c* after 2 msec of folding with protection factors generally less than 5 (for 32 out of 36 residues measured; see Fig 2.3*d*), and interpreted this in terms of the fast formation of the near-native helical content (Sauder & Roder, 1998). A less direct argument is that the burst-phase kinetics appear to be limited by a barrier crossing between two defined states as in Fig. 2.2*c*, insofar as the fast phase (~ 50 μsec) exhibits mono-exponential kinetics, has a significant activation enthalpy ($7.5 \text{ kcal mol}^{-1}$), and its rate is insensitive to initial conditions (Shastry & Roder, 1998). Let us consider the evidence for this view.

In elegant work, Akiyama *et al.* were able to record the entire post-burst-phase CD spectrum of Cyt *c* in fast continuous flow experiments. Fig. 2.3*f* shows the result when folding was initiated by denaturant dilution from 4.4 M to 0.7 M GdmCl (Akiyama *et al.*, 2000). After 400 μsec , the CD of the high GdmCl denatured spectrum changes to a spectrum equivalent with the acid unfolded state (Akiyama *et al.*, 2000), with the thermally unfolded state (Sosnick *et al.*, 1997), with the non-folding Cyt *c* fragments (Sosnick *et al.*, 1997), and also with the species designated as Intermediate I by Akiyama *et al.* The burst-phase simply readjusts the CD to the low denaturant U-state spectrum

common to all of these conditions. Whole spectrum deconvolution (rather than a simple 222 nm ratio) indicates less than 10% helix content in the burst-phase. When refolding was initiated by a pH jump from the acid-unfolded state (in no denaturant), the CD spectrum starts from this U-state condition and there is no burst-phase CD change, even though the FI burst is large. That is, CD recorded in time after the pH jump extrapolates back to the U-state spectrum with no missing amplitude (Akiyama *et al.*, 2000). Widespread helix formation in the burst-phase is also inconsistent with our amide kinetic isotope results (Krantz *et al.*, 2000), which indicate that all of the native hydrogen bonds form later, in the multi-msec folding phase (Fig. 2.3g; see also Ch. 3). These results consistently show that no helical structure forms in the sub-millisecond burst-phase, contrary to the interpretation of fast HX protection with nearly native 40% helical content.

An alternative explanation for the fast HX protection found in Cyt *c* and in other proteins (Hosszu *et al.*, 1997; Parker *et al.*, 1997) considers the impressive sensitivity of HX to local hydrophobic blocking (Bai *et al.*, 1993; Bai *et al.*, 1994; Milne *et al.*, 1998). Bieri and Kiefhaber find generalized low level HX protection ($P_f \sim 5$) in the initially collapsed, but unstructured, form of hen lysozyme (Bieri & Kiefhaber, 2001). Generalized low level HX blocking is especially likely in the concentrated Na_2SO_4 (0.4 M) used in these studies which drives polypeptide chain condensation. Measured HX protection in Cyt *c* may also come from the accumulation of some native protein during the 2 msec pre-HX labeling period, since folding to the native state at the conditions used can proceed in less than 2 msec. It may seem suggestive that the rapidly protected amide hydrogen in Cyt *c* occur only in segments that are protected in the native protein (Fig. 2.3d), but this pattern is imposed by the fact that only these amide hydrogen can be measured.

Let us next consider the evidence favoring fast intermediate formation on the basis of the kinetic character of the burst-phase. Shastry and Roder initiated folding from the acid unfolded state and observed the fast multi-exponential decay of FI due to Förster transfer as Trp59 approaches the heme (see Fig. 2.3e; Shastry & Roder, 1998). Their conclusion—that the burst-phase part of the FI decay exhibits mono-exponential kinetics and therefore signals a distinct barrier crossing event—is difficult to demonstrate. Data recording started after a significant fraction of the overall FI decay had already occurred, and it tracked only the last 40% of the putative mono-exponential burst-phase, from 0.65 to 0.4 in Fig. 2.3e. This phase merges into slower phases fit by additional exponentials. The presence of a large mono-exponential phase seems to be supported by the accurate extrapolation to the initial relative value of 1.0 at zero time (Fig. 2.3e). However, the value of 1.0, measured for the initial U state at pH 2, is artificially low due to acid quenching and also an acid-dependent blue shift in the quenching Soret band toward the tryptophan emission spectrum. The more correct U state FI value is 1.2.

Furthermore, Hagen and Eaton simulated the FI quenching to be expected for a continuous, non-barrier-crossing chain collapse (Hagen & Eaton, 2000; Fig. 2.3e *inset*). Starting from a large Trp to heme distance, the FI decay initially lags and then falls, apparently exponentially, due to the sharp dependence of FI quenching on Trp to heme distance. The simulation is set to a high denaturant value of 0.8 (for 5 M GdmCl). The measured FI decay, however, is normalized to the initial pH 2, no salt condition, where FI is 0.77 of the 5 M GdmCl value (0.8×0.77 of the unquenched 1.0 value in the simulation). Therefore, the burst-phase measured from 0.65 to 0.4 on the FI scale corresponds to 0.40 to 0.25 in the simulation (horizontal lines in Fig. 2.3e). Over this

narrow range of ~ 2 -fold, FI decay can appear to be exponential, and continuous contraction cannot be distinguished from a distinct barrier crossing mechanism.

Other evidence relates to the general character of the FI burst-phase and the chevron roll-over. The burst-phase rate, but not the amplitude, is insensitive to initial conditions (Shastry & Roder, 1998), including pH and the presence of heme ligand (0.2 M imidazole), as might be expected for a barrier crossing $U \rightarrow I$ reaction (Eq. 2.1, Fig. 2.2c). However these same characteristics are also expected for a $U \rightarrow U'$ model (Eq. 2.2, Fig. 2.2d). Similarly, the small activation enthalpy measured for the FI burst-phase (~ 7.5 kcal mol⁻¹) is not diagnostic of barrier crossing, since it is also close to expectations for simple chain contraction, namely 4.2 kcal mol⁻¹ for the temperature dependence of solvent viscosity plus several kcal mol⁻¹ for other contributions including main chain bond rotations (Bolhuis *et al.*, 2000).

It appears that chevron roll-over previously observed for Cyt *c* is due to aggregation rather than intermediate formation, apparently dependent on intermolecular histidine to heme iron ligation. The roll-over in the kinetic chevron plot disappears when Cyt *c* folds at very low protein concentrations (unpublished observations), or with the heme iron protected by bound imidazole (Bhattacharyya & Sosnick, 1999; Chan *et al.*, 1997) or by azide (Krantz *et al.*, 2000), or in a mutant with the two peripheral histidines replaced (Fig. 2.3g). Under all of these conditions, roll-over is not seen even at very low denaturant where large burst-phase FI loss occurs and the putative burst intermediate was postulated to go from minimally to highly populated (Sauder & Roder, 1998). The absence of chevron roll-over is inconsistent with the formation of a well-structured

intermediate, because the burial of an appreciable amount of surface before the rate-limiting step must reduce the slope of the chevron folding arm. Notably, a large burst FI signal occurs also in reduced Cyt *c* where folding was demonstrated to be two-state (Bhuyan & Udgaonkar, 2001). These results highlight the danger of inferring global behavior from a single (maximally hydrophobic) indole group.

In other work, Hagen and Eaton were able to measure fast temperature jump-induced FI changes in Cyt *c* under certain limited conditions (1.2 to 2.2 M GdmCl), where the U state is marginally populated (Hagen & Eaton, 2000). They observed a multi-phase process with the fastest rate, roughly exponential, about 4-fold slower than the burst-phase rate fit by Shastry and Roder. Hagen and Eaton considered the fast exponential phase to reflect a barrier crossing re-equilibration between minimally populated compact and extended denatured states, as in the earlier fast FI study (Shastry & Roder, 1998). An interesting alternative can now be considered. Under the limited conditions found usable by Hagen and Eaton, three folding intermediates defined by native state HX (Bai *et al.*, 1995a) are present at a higher level than U (Mayne & Englander, 2000) and are very close to each other in free energy. This suggests that the temperature jump experiment may have measured the redistribution of pre-existing populations of partially folded Cyt *c* intermediates.

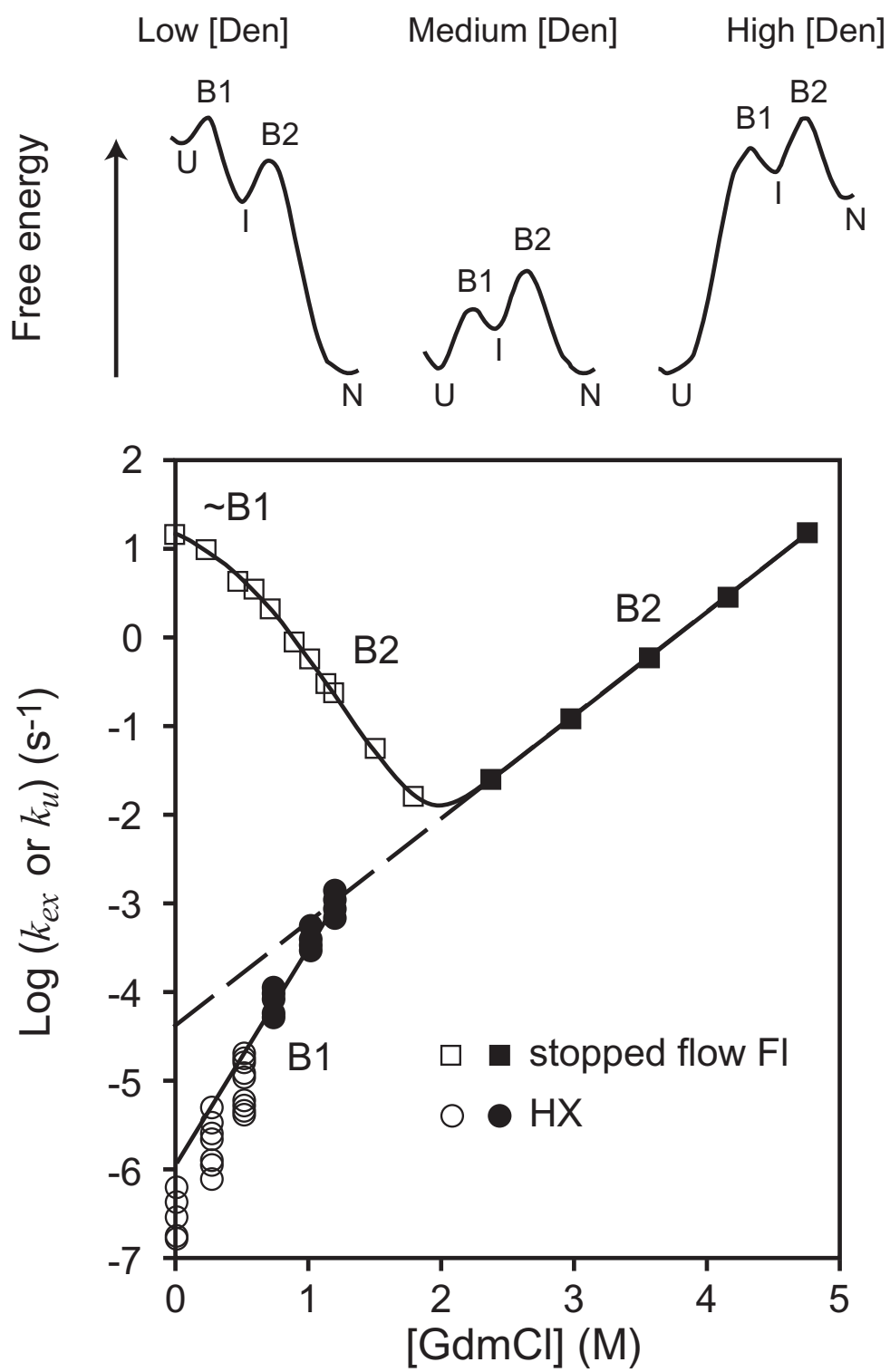
In summary, a great deal of available information bears on the structure reached in the Cyt *c* burst-phase. It exhibits the same FI and CD as the acid and thermally denatured states and non-folding fragment models, with less than 10% helix. The native H-bonds form much later (H/D isotope effect; Fig. 2.3g; see also Ch. 3). It has a minimal amount of buried surface (no roll-over under conditions that avoid aggregation). Supposedly, mono-

exponential kinetics suggested to require a distinct barrier-crossing transition do not do so. Other results suggested to support a structured intermediate—e.g. temperature dependence, minor HX protection, independence of rate on initial conditions—seem equally consistent with a fast solvent-dependent U to U' chain contraction. These and other considerations are against the formation of a distinct structured burst-phase intermediate in Cyt *c* folding (Eq. 2.1), but they are all fully consistent with a fast solvent-dependent readjustment of the still unfolded U state (Eq. 2.2).

2.3.2 Barnase

Barnase has been intensively studied by Fersht and co-workers as a paradigm for three-state folding. Both burst-phase and later intermediates have been thought to accumulate. Initial HX pulse labeling results (Bycroft *et al.*, 1990; Matouschek *et al.*, 1992) were interpreted to indicate the accumulation of an early burst-phase intermediate, a later intermediate (~70 msec) with the secondary elements formed, and a 200 msec final folding phase. The initial sub-millisecond accumulation of a folding intermediate was supported by a missing burst-phase amplitude (FI and CD), curvature in the folding chevron (Fig. 2.4), and the failure of the chevron to obey two-state criteria ($-RT \ln k_u/k_f = \Delta G_u^o$; $m_u - m_f = m^o$; see App. 3). The free energy deficit between the equilibrium stability of native barnase and its stability determined kinetically (from k_u and k_f extrapolated to zero denaturant), led to an estimate of 2 to 4 kcal mol⁻¹ in stability for the burst-phase intermediate in wild type barnase and various mutants. These observations turn out to be misleading. Early HX pulse labeling results are not reproducible (Chu *et al.*, 1999), because they apparently depend on aggregation, attributed to the high protein concentration used (Fersht, 2000a) and to a contaminant (Takei *et al.*, 2000).

Figure 2.4 Barnase folding and unfolding. Unfolding rates at low denaturant, obtained from native state HX data (Clarke & Fersht, 1996) under EX1 and near EX1 conditions for a number of amide hydrogen that exchange by way of global unfolding (\bigcirc), curve down (*solid line at low denaturant*) below the extrapolated on-chevron unfolding behavior, accounting for the missing stabilization free energy initially assigned to an early burst-phase intermediate (Takei *et al.*, 2000). The pathway free energy profiles (*upper*) explain the chevron curvature in terms of the denaturant-dependent crossing over of barriers (B1 and B2), which requires an on-pathway intermediate that does not accumulate (Fersht, 2000a).



Burst-phase changes and chevron curvature are now known to have other interpretations, including solvent-dependent condensation of the U state ensemble (Eq. 2.2) (Parker & Marqusee, 1999; Qi *et al.*, 1998; Sosnick *et al.*, 1997), aggregation (Silow & Oliveberg, 1997), and movement of the transition state barrier (Matouschek *et al.*, 1995; Oliveberg *et al.*, 1998; Otzen *et al.*, 1999). Melting analysis applied to the burst-phase amplitude as a function of urea concentration (CD and FI) showed that the state reached in the burst-phase has stability close to zero relative to U (Takei *et al.*, 2000). HX competition results revealed protection factors of two or less indicating that no significantly protected intermediate accumulates before final folding (Takei *et al.*, 2000). The missing 3 kcal mol⁻¹ of wild type barnase stability, initially assigned to the putative intermediate, is accounted for by the fact that unfolding rates curve sharply downward at low denaturant, well below the value linearly extrapolated from the chevron (Fig. 2.4). This was shown (Takei *et al.*, 2000) by analyzing data for unfolding rates at low denaturant from earlier native state HX measurements under EX1 conditions (Clarke & Fersht, 1996).

In response, Fersht (Fersht, 2000a) does not contest the absence of a populated intermediate (burst-phase or otherwise) during refolding experiments. However, he lists a number of observations that favor the situation diagrammed in Fig. 2.4 (*top*) with at least one on-pathway intermediate. He especially stresses the sharpness of the downward curve of the chevron unfolding arm at low denaturant, which suggests a crossing over of two sequential barriers as denaturant decreases, indicating the presence of an on-pathway intermediate. Fersht suggests that the aggregation that occurred in the earlier HX pulse labeling study may have acted fortuitously to stabilize this real intermediate which otherwise does not accumulate.

In summary, barnase obeys the thesis discussed here. In folding experiments initiated by denaturant dilution, unfolded barnase produces burst-phase signals and chevron curvature in response to the solvent change, but it forms no stable structure that is detectably different from U. After an initial rate-limiting folding step (~50 msec), folding may proceed through one or more intermediates (Fig. 2.4) but they do not accumulate. The analyses presented by both Bai (Takei *et al.*, 2000) and Fersht (Fersht, 2000a) agree with this view.

2.3.3 Ubiquitin

Previously, Khorasanizadeh *et al.* engineered a single tryptophan (F45W) into the core of ubiquitin (Ub) to provide a F1 probe and used it to measure the folding of this protein and some mutants as a function of GdmCl concentration (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996). In low denaturant, folding at 25°C but not 8°C exhibited a burst-phase loss in F1 amplitude and a roll-over in the folding chevron, suggesting a stable burst-phase intermediate.

On the other hand, a variety of other results for Ub are consistent with a two-state folding behavior. Hydrogen exchange (HX) labeling studies (Briggs & Roder, 1992; Gladwin & Evans, 1996; Khorasanizadeh *et al.*, 1993) failed to observe amide H-bond formation in the first few milliseconds of refolding. In fact, studies by Gladwin and Evans at 20°C indicated that all the amide protons exchange with protection factors near unity indicating that the backbone is as exposed as in random coil polypeptide (Gladwin & Evans, 1996). Likewise, their stopped-flow far-UV circular dichroism studies failed to

observe any burst-phase intermediate (Gladwin & Evans, 1996). These data argue against the proposed kinetic intermediate (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996), because any such species having the majority of the total surface buried in the U→N reaction should have had some secondary structure and HX slowing.

Furthermore, Ub folds in a two-state manner at lower temperatures, as seen at 8°C (Khorasanizadeh *et al.*, 1993), and 10°C (Krantz *et al.*, 2000), where folding rate constants are readily measurable with a conventional stopped-flow apparatus. In these studies, the folding rate constant remains linear over all denaturant concentrations and the fluorescent signal is completely accounted for in the observed phase. Also, our D/H amide kinetic isotope studies (see Ch. 3) suggest only minimal structure formation is present prior to the transition state (Krantz *et al.*, 2000).

Lastly, sub-microsecond temperature-induced refolding measurements by Gruebele and coworkers on cold denatured ubiquitin show no evidence of an intermediate at 2°C after accounting for a solvent-dependent response of the unfolded state (Sabelko *et al.*, 1999). More complex kinetics are observed at 8°C, but the reaction surface still is viewed as consistent with a two-state system. The very fast folding population represents a sub-fraction of the -8°C, cold-denatured ensemble whose configuration is such that upon the nanosecond temperature-jump, these molecules reside on the native-side of the rate-limiting barrier on the new +8°C reaction surface.

Given these apparently conflicting results for such a well-studied protein and the importance of understanding the role of stable kinetic intermediates, we investigated whether Ub's folding behavior is kinetically two- or three-state. Specifically, we

examined whether a two-state model is adequate to account for the denaturant dependence of folding rate constants and the amplitude of the observed signal changes. We have found that a two-state model adequately describes the folding pathway of Ub under conditions where an intermediate reportedly existed (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996). Specifically, we modified our instrument to reduce the dead time to one millisecond and increased the signal-to-noise (App. 4). Also, we used continuous-flow and double-jump stopped-flow protocols to minimize slow folding phases due to improper proline isomerization and other effects.

2.3.3.1 *Single-jump measurements*

We performed stopped-flow fluorescence measurements in 20 mM sodium acetate, pH 5, 25 °C (Fig 2.5a). As observed in previous refolding studies of Ub at this condition (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996), the fluorescence trace is largely, but not completely monophasic, with the major phase representing folding to the native state (Khorasanizadeh *et al.*, 1993). As with certain other proteins including chymotrypsin inhibitor 2 (Jackson & Fersht, 1991b) and *src*-SH3 domain (Grantcharova & Baker, 1997), the minor slow phases in Ub are predominantly associated with a population of slow folding molecules containing non-native *cis*-proline isomers (native Ub has three *trans*-prolines).

Figure 2.5 Ubiquitin folding traces. Denaturant dependence of folding at pH 5.0, 25 °C.

Raw data traces in **a**, single-jump refolding experiments and **b**, double-jump refolding experiments which eliminate extraneous proline phases. The non-coincidence of the fully folded values largely represents the denaturant dependence of the native state (see Fig. 2.6b). *Inset:* total amplitude change. In the two-state model, the change should be constant given that the slope of the denaturant dependent base-lines of the denatured and native states are equivalent, as is the case (Fig. 2.6b). The data, sampled every 50 μ sec, are binned into increasing longer time intervals which results in the decrease in the noise level.

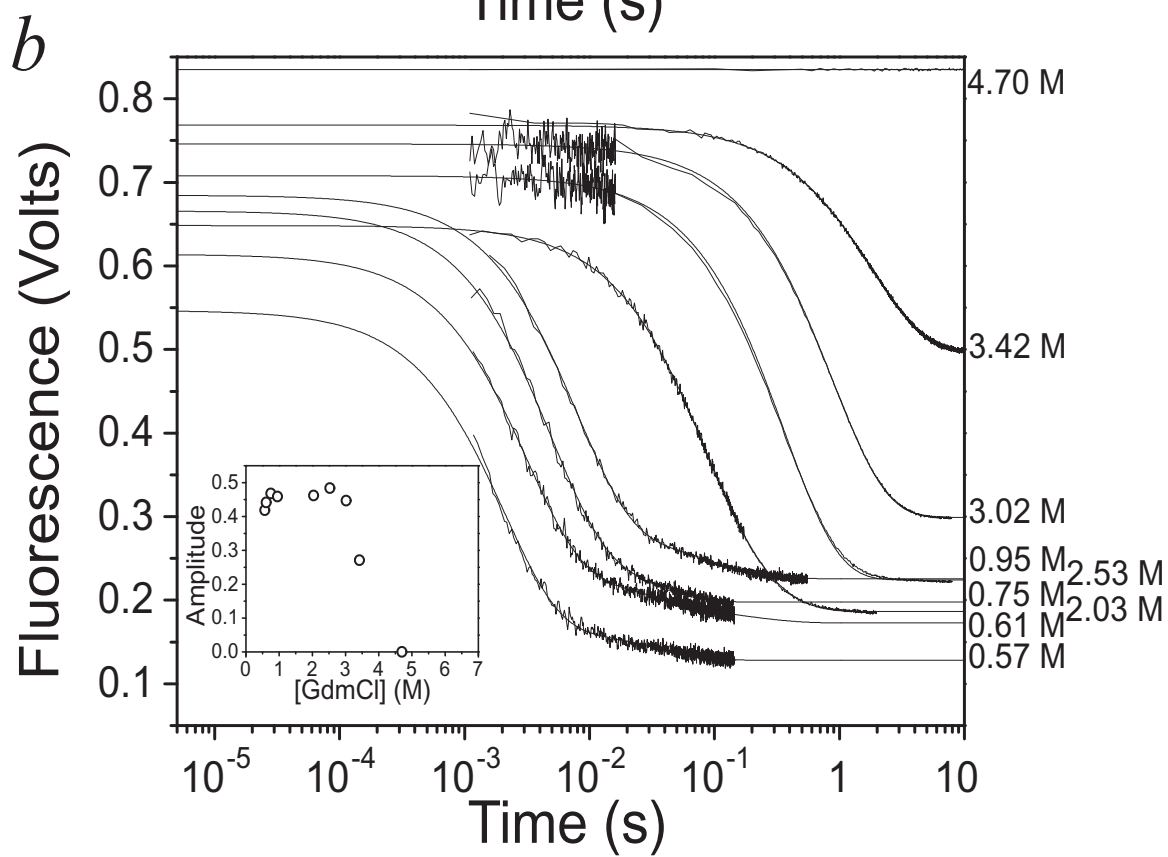
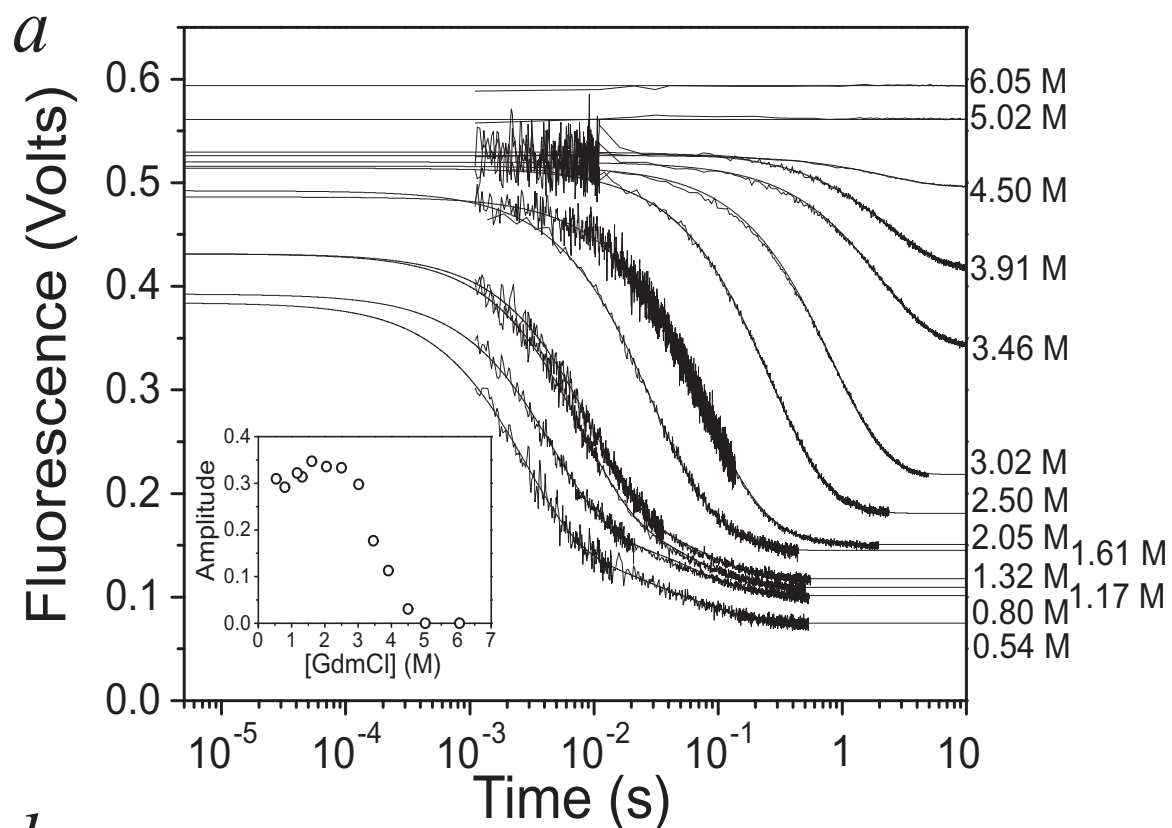


Table 2.1 Thermodynamic and kinetic parameters for ubiquitin folding

Condition	$-\Delta G_{equil}$ (kcal mol ⁻¹)	m_{equil}^o (kcal mol ⁻¹ M ⁻¹)	$-\Delta G_{kinetics}$ (kcal mol ⁻¹)	$RT \ln k_f^{H_2O}$ (kcal mol ⁻¹) ^c	$m_{kinetics}^o$ (kcal mol ⁻¹ M ⁻¹)	m_f (kcal mol ⁻¹ M ⁻¹)
25 °C ^a	8.19±0.10	2.18±0.02	8.40±0.24	4.35±0.04	2.17±0.04	1.35±0.02
25 °C / 0.23 M Sulfate ^a	9.50±0.37	2.13±0.09	10.2±0.5	5.18±0.06	2.23±0.09	1.31±0.03
5 °C / 1.0 M NaCl ^b	8.42±0.25	2.23±0.06	8.20±0.20	3.73±0.03	2.18±0.04	1.37±0.02
15 °C / 1.0 M NaCl ^b	8.71±0.22	2.24±0.05	9.28±0.27	4.49±0.03	2.24±0.05	1.32±0.02
25 °C / 1.0 M NaCl ^b	9.39±0.37	2.23±0.09	9.70±0.34	5.11±0.03	2.28±0.06	1.32±0.02

^a20 mM Sodium Acetate, pH 5.0.

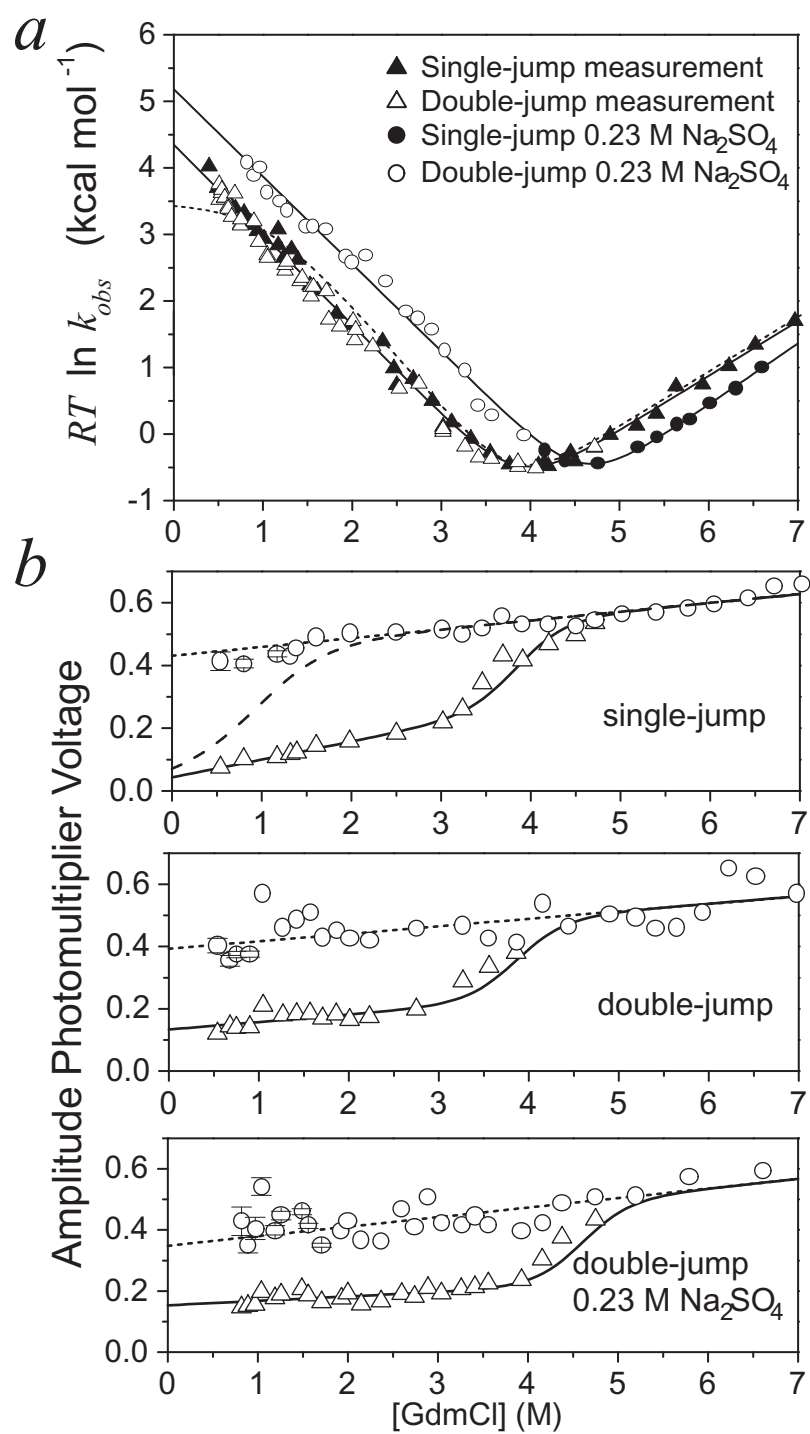
^b20 mM Sodium Acetate, pH 4.5.

^cExtrapolated refolding activation energy in the absence of denaturant.

Our measured folding and unfolding activation free energies for the major phase are in very good agreement with the earlier studies (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996) except that our folding values remain linear down to the lowest denaturant concentration measured (0.5 M GdmCl) and the resulting chevron plot has the ideal \vee -shape typical of a two-state folding proteins (Fig. 2.6a). Although the rate constants approach the limit of the stopped-flow apparatus, $\sim 50\%$ of the amplitude is observable for a rate constant of 500 s^{-1} . Additionally, the entire fluorescence signal is accounted for in the observed phase upon extrapolation to zero time (Fig. 2.5, *inset* and Fig. 2.6). Neither of the two signatures of a folding intermediate, a saturation in refolding rate constant and a burst-phase amplitude, are observed at low denaturant concentrations. Furthermore, the kinetic data interpreted in the context of a two-state model reproduce the equilibrium stability and surface area burial as determined from independent equilibrium denaturation measurements (Table 2.1).

It should be noted that an accurate determination of the fastest rate constants is difficult, because they are near the limit of our apparatus and folding is heterogeneous, in large part due to proline isomerization. The fluorescence traces are multi-phasic and three to five exponential terms are required to fit the data over an extended time range. The second fastest phase accounts for up to $\sim 20\%$ of the amplitude and its rate constant is within a factor of five of the fastest rate constant. The rate constants shown in Fig. 2.6a are obtained using a bi-exponential fit over a limited time range at GdmCl concentrations below 2 M and mono-exponential fits above 2 M.

Figure 2.6 Ubiquitin denaturant dependence. Denaturant dependence of folding at pH 5.0, 25 °C. **a**, Folding activation energies for single and double-jump protocols in the absence and presence of 0.23 M sodium sulfate. A previous fit (.....) to Ub F45W kinetic data is given (from Khorasanizadeh *et al.*, 1996). **b**, The raw fluorescence signal (amplitude photomultiplier voltage) for single-jump, and for double-jump measurements in the absence and presence of 0.23 M sodium sulfate. Amplitudes are defined by the raw fluorescence signals at the beginning (\bigcirc) after extrapolation to $t=0$ and end (\triangle) of each refolding kinetic measurement, where the latter are fit (—) to a two-state equilibrium denaturation profile using values from Table 2.1. A linear fit (.....) to the denaturant dependence of the unfolded state's fluorescence signal is also provided. The sigmoidal curve (--) in the top single-jump panel is the fit to the burst-phase amplitudes (from Khorasanizadeh *et al.*, 1996). Note how the present data neither reflect the loss of signal in the burst-phase (**b**) nor roll-over in the observed kinetics (**a**). Error bars at low denaturant concentrations represent the uncertainty of the extrapolated amplitude when the folding half-life is on the order of the instrumental dead-time.



The present combination of fast folding and heterogeneity can result in the fastest rate constants being underestimated. A portion of the fluorescence signal is unobservable due to the dead-time of the stopped-flow apparatus. As the folding rate constant increases, less of the major phase is observed and the proline-related phases contribute progressively more to the measured portion of the signal. These proline-related phases then may inadvertently be included in the fit to the fastest phase. Hence, as the rate of the fastest phase increases, the true rate constant can be increasingly underestimated. This explanation may account for some of the difference between the present and the previous studies (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996). The dead-time, τ_d , in the previous studies was longer, about 2.2 ms (Khorasanizadeh *et al.*, 1993), which would have exacerbated this effect as well.

Furthermore, an underestimation of the folding rate constant will produce an apparent burst-phase amplitude even when one does not actually exist. The zero time amplitude of the measured phase is calculated based upon the folding rate constant: $A_{observed}(t_0) = A_{observed}(\tau_d)e^{k \times \tau_d}$. An underestimated rate will, upon extrapolation back to t_0 , reduce the reported amplitude of the major phase, $A_{observed}(t_0)$, thereby generating an apparent missing amplitude. Therefore, both a burst-phase loss of amplitude and non-linear chevron behavior can come from a single cause, an underestimation of folding rate constants. They do not necessarily provide independent evidence for the formation of a kinetic intermediate.

2.3.3.2 Double-jump measurements

Given the difficulty of measuring rate constants in a heterogeneous folding mixture, it is desirable to decrease the amplitude of the minor phases. Ubiquitin has three *trans*-prolines, residues whose isomerization in the unfolded state is known to produce a slow-folding population (Khorasanizadeh *et al.*, 1993). We implemented a double-jump ($N \rightarrow U \rightarrow N$) refolding protocol to reduce the proline-related slow phases. The native protein is allowed to unfold at 4.7 M GdmCl for ten seconds (about ten unfolding half-lives). Before the prolines in the native, *trans* configuration have a chance to isomerize in the denatured state, the concentrated denaturant solution is diluted and folding proceeds. This procedure produces a denatured population that, unlike the single-jump population, contains essentially only the native proline configurations. The result of this protocol is a decrease in the amplitude of the slower phases that also are better resolved from the fastest phase (Fig. 2.5*b*). The rate constants of the fastest phase measured in this manner are in excellent agreement with those from the single-jump measurements (Fig. 2.5*a*). There is no indication of a plateau in the refolding rate constant, and the zero-time value of the fluorescence signal does not have a burst-phase loss of amplitude (Fig. 2.6*b*).

2.3.3.3 Continuous-flow measurements

Given the discrepancy between the present and previous folding studies (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996), we sought to examine whether instrumental artifacts could exist in our measurements (e.g. inaccurate dead-time measurement or inadequate mixing). We conducted multiple continuous-flow measurements at different flow-speeds to alter the dead-time, or the unobservable portion

of the reaction (Fig. 2.7). The fluorescent level is measured for about 20 msec for reactions having aged for times between 1.3 and 11 msec. The values for each continuous-flow measurement represent single time points in the refolding experiment and provide information on the refolding rate constant as well as the fluorescence amplitude. The slope the natural log of the amplitude change versus τ_d plot is the refolding rate constant. Likewise, the y-intercept is the zero-time fluorescence signal.

Measurements are performed at four different denaturant concentrations between 0.5 and 2.1 M GdmCl. These denaturant concentrations are chosen as they span the range where a saturation in rate constants and a burst amplitude was reported (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996). The refolding rate constants determined from the slope of the plot (Fig. 2.7) agree nearly within statistical error with those obtained in the standard stopped-flow measurements (Table 2.2). Furthermore, the total fluorescence change (y-intercept) is identical for refolding at all four GdmCl concentrations. The self-consistency between these and the conventional stopped-flow measurements eliminate inaccurate dead-time values or mixing artifacts as sources for the discrepancy between the present and previous studies. We believe this comparison between stopped-flow and continuous-flow measurements *conducted under the actual refolding conditions* is one of the most accurate methods to confirm the instrument's calibration, because the dead-volume and flow-speed are corroborated by the highly accurate timing of the A/D board. Most importantly, the measurements are independent demonstrations that refolding rate constants do not plateau and all of the fluorescence signal is accounted for in the observed phases.

Figure 2.7 Ubiquitin continuous-flow measurements. Continuous-flow values are obtained by varying the flow rate of the peptide and buffer solutions. Because the flow rate is inversely proportional to the dead-time (τ_d) of the instrument, differently aged protein refolding solutions can be produced as indicated by the time axis. The four denaturant conditions, 0.54 M (●), 0.80 M (■), 1.32 M (▲), and 2.05 M (▼), are chosen to test for the presence of a burst-phase as well as to demonstrate that the instrument's flow speed is calibrated. The common intercept indicates that all of the change in the raw fluorescence signal, A_{obs} , is accounted for in the observed phase regardless of denaturant concentration, and no submillisecond process occurs in the instrumental dead-time. Open symbols at $\tau_d = 0$ represent the natural logarithm of the voltage change determined from standard stopped-flow measurements with a 1.3 ms dead-time for 0.54 M (○), 0.80 M (□), 1.32 M (△), and 2.05 M (▽). A comparison of the amplitude change and the rate constants (i.e. the slope) to traditional stopped-flow measurements are presented in Table 2.2.

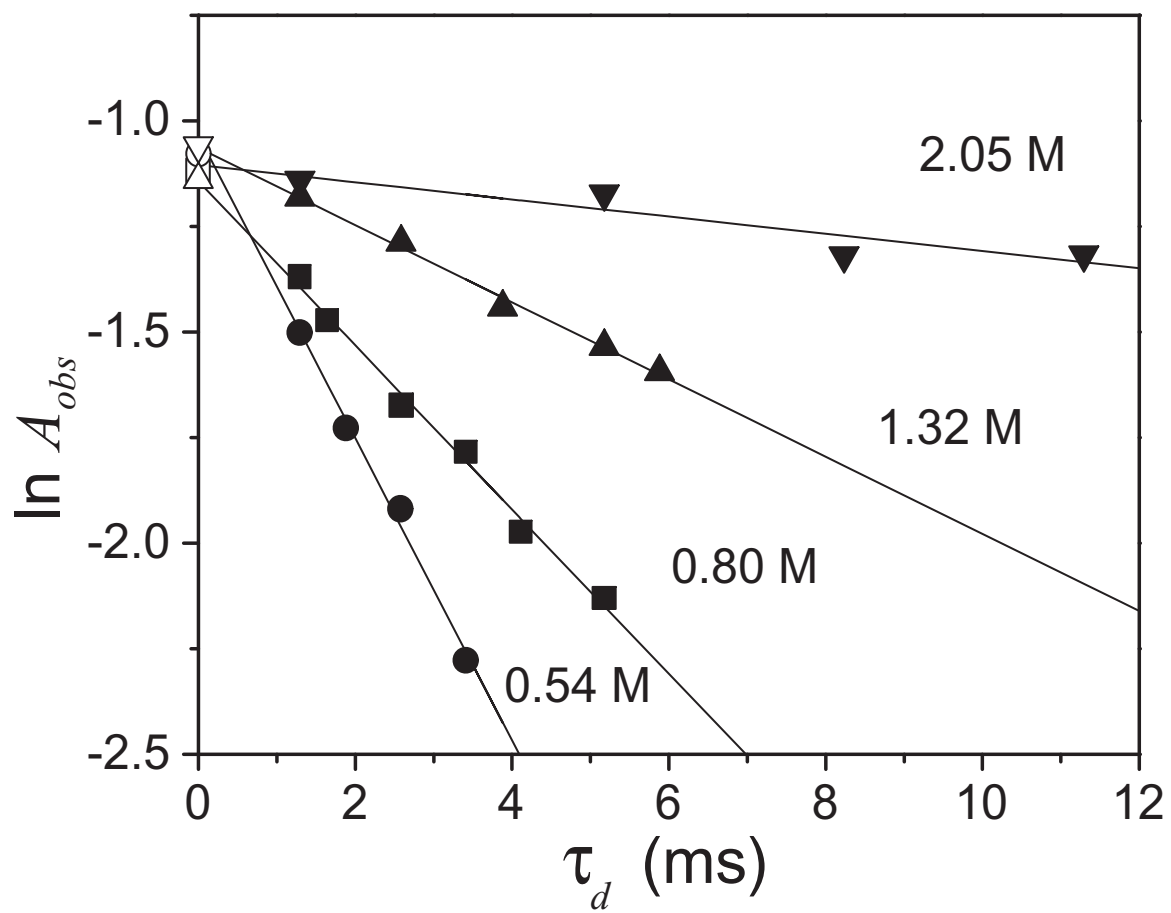


Table 2.2 Stopped and continuous-flow measurements for ubiquitin

[GdmCl] (M)	Stopped- flow rate constant ^a (s ⁻¹)	Continuous- flow rate constant ^b (s ⁻¹)	Stopped-flow amplitude ^a (Volts)	Continuous- flow amplitude ^b (Volts)
0.54	368±15	356±19	0.34±0.03	0.36±0.05
0.80	213±2	194±9	0.31±0.02	0.32±0.03
1.32	107±1	92±8	0.32±0.01	0.35±0.04
2.05	16.2±0.5	20.3±4	0.35±0.01	0.33±0.03

^aSingle-jump refolding traces with a 1.3 ms dead-time.

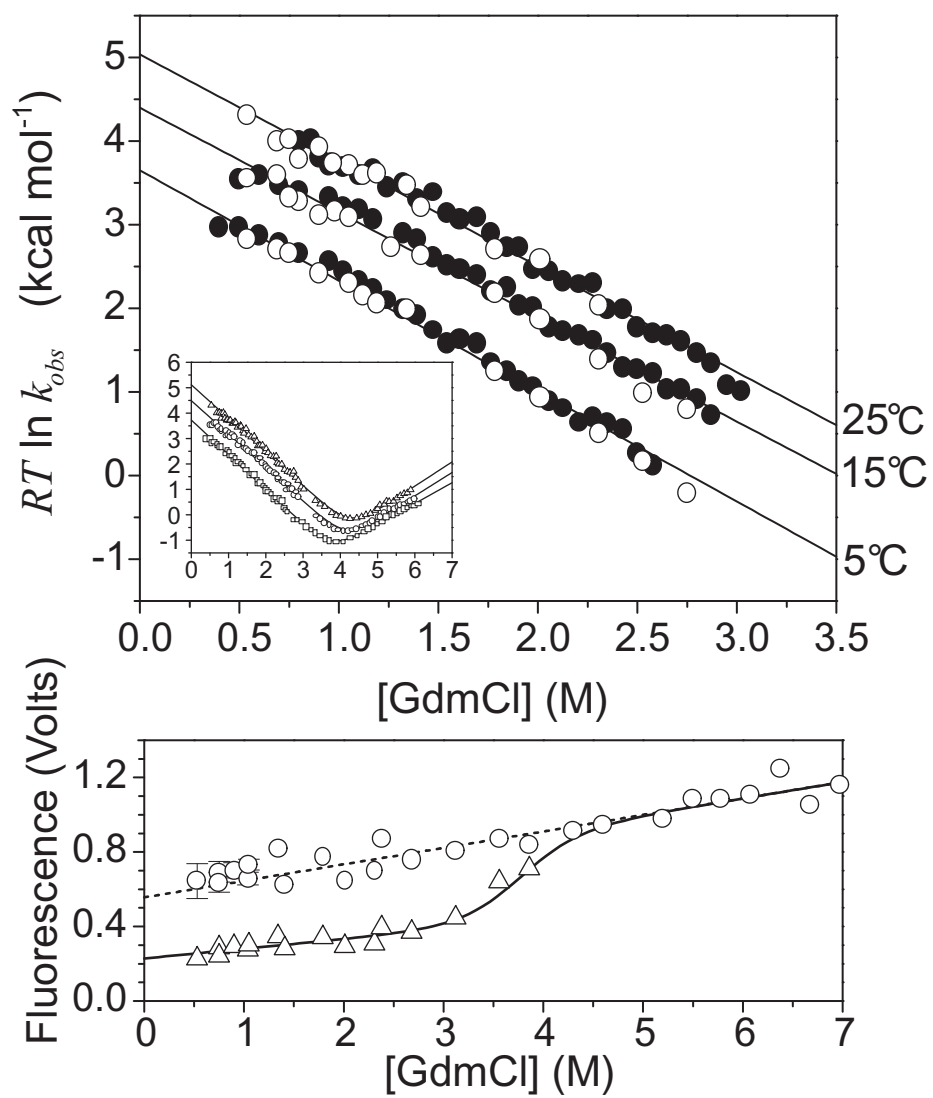
^bFrom the slope and y-intercept of the plot in Fig. 2.7.

2.3.3.4 Stabilizing additives

Previous measurements indicated that the presence of 0.4 M sodium sulfate increases the stability of the proposed kinetic intermediate (Khorasanizadeh *et al.*, 1996). Folding rate constants at 25 °C in the presence of 0.4 M sodium sulfate, however, are at the measurement limit of our apparatus, and we would not be able to confidently distinguish between two- and three-state folding models. Hence, we have measured folding in the presence of a lower level of sulfate, 0.23 M, where a potential deviation from two-state behavior would be readily observable. To increase the accuracy of these measurements, they are conducted in double-jump mode.

The presence of sodium sulfate results in a rightward translation of the chevron (Fig. 2.6a) with the folding and unfolding activation energies changing by about 1 and $-0.5 \text{ kcal mol}^{-1}$, respectively. As in the absence of additives, all of the fluorescence signal is fully accounted for in the observed phases (Fig. 2.6b) and no roll-over in rate constants is observed in spite of a potential 1 kcal mol^{-1} stabilization of a intermediate. Sodium sulfate is commonly employed as a protein precipitate and can lead to aggregation. Because our studies and previous studies (Khorasanizadeh *et al.*, 1996) revealed an additional slow phase in the presence of sulfate, we sought an alternative stabilizing additive. Work by Makhatadze and co-workers indicated that the many types of anions stabilize ubiquitin (Makhatadze *et al.*, 1998). We chose 1.0 M sodium chloride, because chloride ions are already present in the GdmCl buffer. While not as stabilizing as sodium sulfate ($\Delta\Delta G=1.8$ versus 1.3 kcal M^{-1}), sodium chloride largely reduced the sulfate-dependent slow phase. In the high salt condition, folding rate constants do not

Figure 2.8 Ubiquitin folding in the presence of a stabilizing additive. Folding in 1 M sodium chloride, at pH 4.5. **(a)** folding activation energy for single-jump (●) and double-jump (○) measurements at 5, 15 and 25 °C with corresponding refolding activation energy versus GdmCl chevron plots (*inset*). **(b)** Fluorescence amplitudes at 25 °C are defined by the raw fluorescence signals at the beginning (○) after extrapolation to $t=0$ and end (△) of each refolding kinetic measurement. The solid line is a fit to the equilibrium denaturation profile using values from Table 2.1. The dashed line is the denaturant dependent baseline of the unfolded state.



plateau, and all the fluorescence signal is accounted for in the observed phase (Fig 2.8). As with the previous measurements, a two-state model is adequate to describe Ub folding, even in the presence of stabilizing additives.

2.3.3.5 Importance of stopped-flow dead time

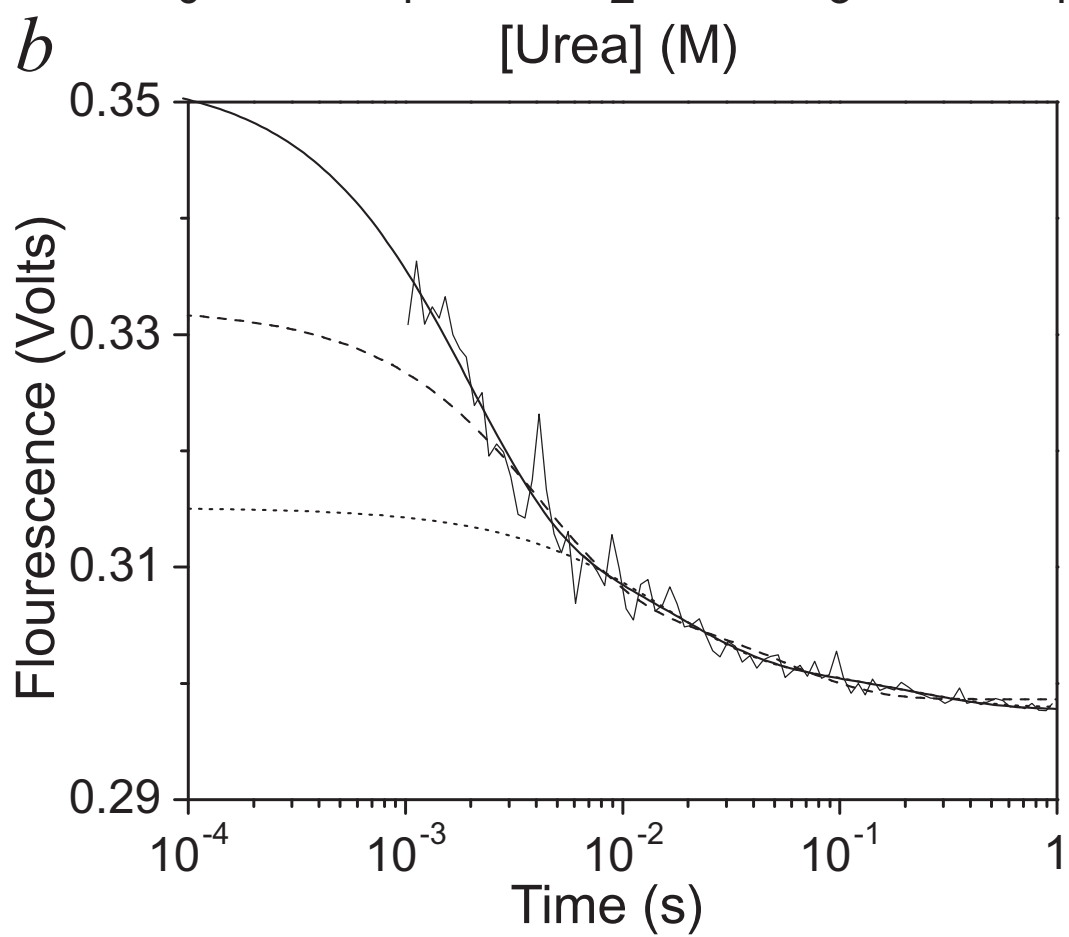
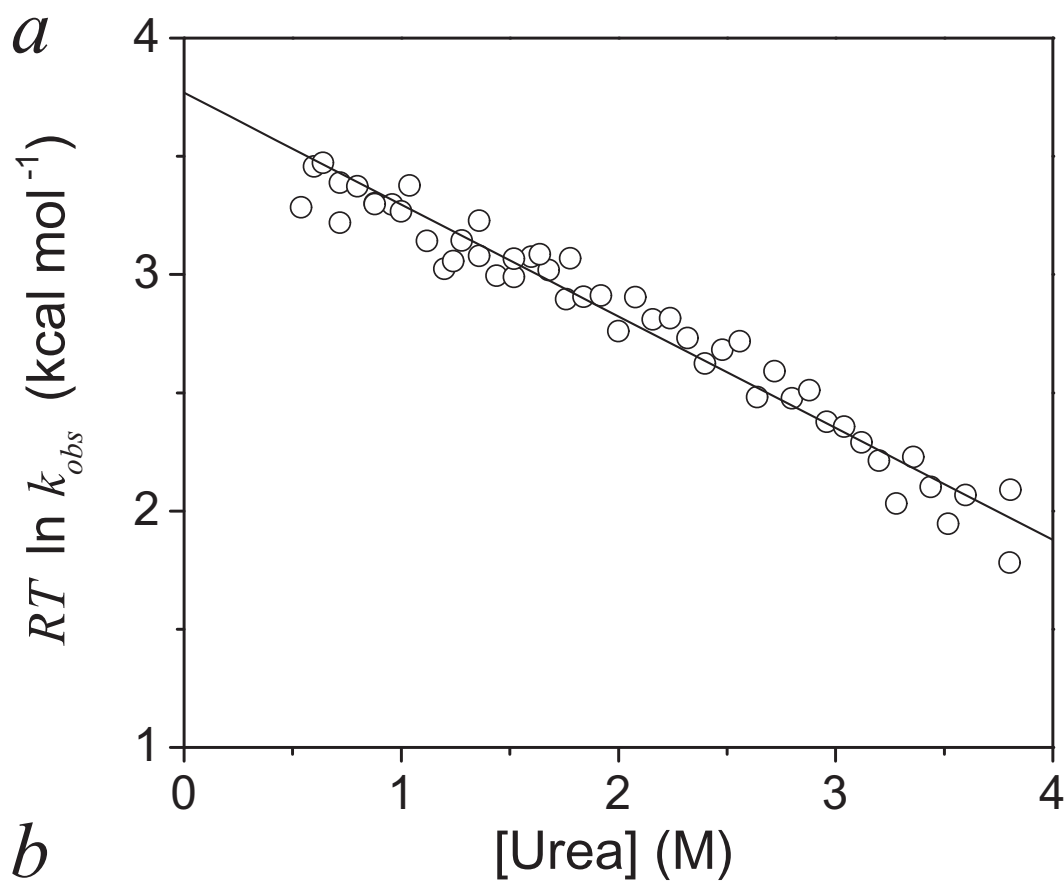
A stopped-flow dead time of ~ 2 millisecond is only adequate in distinguishing rates up to $\sim 500 \text{ s}^{-1}$ when the kinetics are pure mono-exponentials. If, as in the case of ubiquitin, a slow phase appears on the order of $30\text{-}40 \text{ s}^{-1}$, then there is significant risk that the slower rate will be over-represented in the curve fit due to the fact that the stopped-flows dead-time limits the amount of the fast rate on the order of 300 to 400 s^{-1} that can be observed. I specifically modified our stopped-flow to improve the signal-to-noise and dead-time to 1 millisecond (App. 4), and determined that slower stopped-flows (2-5 msec) can lead to a false burst-phase amplitude. This effect is demonstrated by fitting a sample refolding trace for ubiquitin in urea denaturant (Fig. 2.9). Finally, note that even in the lowest concentrations of urea, ubiquitin fails to exhibit roll-over.

2.3.3.6 Discussion of ubiquitin

Given the strong constraint two-state folding behavior places on possible folding models, it is important to establish when folding deviates from a two-state model. The most frequently applied method to establish two-state behavior is the "chevron analysis." When all the standard free energy and surface area burial are accounted for in the observed single exponential reaction, a burst-phase intermediate cannot exist which is stable relative to the denatured state and buries a significant amount of surface area.

Figure 2.9 Dead time and the accurate determination of the folding rate of ubiquitin.

Refolding activation energy as a function of urea concentration for 0.5 μM Ub F45W in the presence of 0.4 M Na_2SO_4 , 20 mM acetate, pH 5.0, 10 $^\circ\text{C}$ using fluorescence **(a)**. The linear fit parameters to the data are $RT \ln k_f^{\text{H}_2\text{O}} = 3.77 \pm 0.04 \text{ kcal mol}^{-1}$, and $m_f = 0.47 \pm 0.02 \text{ kcal mol}^{-1} \text{ M}^{-1}$. An example kinetic trace from stopped flow fluorescence (1 ms dead time) is fit with two exponential decays (**—**) where the fast rate (400 s^{-1}) is plotted above **(b)**. The slow rate (34 s^{-1}) presumably is due to either proline isomerization or transient protein aggregation. This slower phase contaminates mono-exponential fits using a 2.2 (**--**) or 5 ms (**....**) dead-time, as demonstrated by the fits shown, thereby incorrectly resulting in chevron "roll-over". The underestimation of refolding rate erroneously results in a burst-phase amplitude when one does not exist.



Conversely, when a protein folds through a stable intermediate, a decrease in the denaturant dependence of refolding rate constants is observed (Fig. 2.1b) (Khorasanizadeh *et al.*, 1996). Such “roll-over” behavior is seen in medium-sized proteins such as RNase H where the formation of a specific intermediate is confirmed by a variety of other methods including HX pulse labeling (Raschke & Marqusee, 1997), mutation (Raschke *et al.*, 1999), and native-state HX (Chamberlain *et al.*, 1996). In some other systems, however, such roll-over behavior is attributed to protein aggregation (Silow *et al.*, 1999), movement of the transition state towards the unfolded state (Otzen *et al.*, 1999) or inadequate pH control upon buffer dilution. Some other burst-phase signals may be due to the denatured polypeptide responding to the new, poorer solvent conditions rather than to formation of a distinct folding intermediate (Chan *et al.*, 1997; Qi *et al.*, 1998; Sosnick *et al.*, 1996b; Sosnick *et al.*, 1997), although other interpretations have been proposed (Akiyama *et al.*, 2000; Hagen & Eaton, 2000; Shastry & Roder, 1998).

Ub is a model protein for which conflicting data exist regarding whether it folds via a stable kinetic intermediate. In the present studies, we did not detect the previously observed signatures for such a species, either a plateau in refolding rate constants or a burst-phase loss of signal (Khorasanizadeh *et al.*, 1993; Khorasanizadeh *et al.*, 1996). Our observation is seen both in the presence and absence of stabilizing additives and in single-jump, double-jump and continuous-flow modes. In conjunction with previous HX pulse labeling (Briggs & Roder, 1992; Gladwin & Evans, 1996; Khorasanizadeh *et al.*, 1993) and stopped-flow circular dichroism studies (Gladwin & Evans, 1996) that failed to observe H-bond and secondary structure formation in the first few milliseconds of

folding, we conclude that a two-state folding model is adequate to describe Ub folding. Just as importantly, these studies stress that additional experimental tests, including measurements of the protection from HX and the application of double-jump techniques, are extremely useful in establishing when folding rate constants deviate from two-state behavior and a folding intermediate truly exists.

2.3.4 Protein G, B1 domain

Structural information on fast protein G (ProtG) folding was sought by Kuszewski *et al.* (Kuszewski *et al.*, 1994) in an HX competition labeling experiment, with folding initiated by a pH jump (zero denaturant). Calculated HX protection factors in the fast phase, when corrected using the intrinsic free peptide HX rates (Bai *et al.*, 1993), are almost all below 3.5 for 24 out of 26 residues measured. The core hydrophobic residues appear to be the least protected and the polar surface residues most, which seems inconsistent with the formation of a structured intermediate. The low protection factors show a strong inverse correlation ($R = -0.83$) with the intrinsic unprotected amide HX rates (Bai *et al.*, 1993), suggesting a small systematic error, perhaps in the normalizing NMR crosspeak amplitude, or due to faster folding than expected, with true protection factors very close to 1.

Park *et al.* (Park *et al.*, 1999) followed the folding of ProtG in 0.4 M Na₂SO₄ by fast continuous-flow (0.2 to 1.3 msec) and slower stopped-flow (>3 msec) methods, using a single tryptophan F1 probe. When the non-overlapping kinetic data sets were merged,

Figure 2.10 Mono-exponential two-state folding of Protein G. Kinetic traces, amplitudes, and chevrons for ProtG were obtained at the same conditions used previously (Park *et al.*, 1999), 0.4 M Na₂SO₄, 20 mM acetate, pH 5.0, 20-30 μ M protein, at 20°C (**a,b**) and 10°C (**c,d**). The quality of the mono-exponential fits is illustrated by residuals shown for some of the kinetic traces (**a,c lower**). All results consistently confirm 2-state folding, as follows. First, chevron plots show no roll-over. Second, total measured FI amplitudes from the kinetic traces show the absence of denaturant-dependent burst-phase loss (**b,d lower**). Third, two-state chevron criteria are rigorously obeyed, as follows: At 10°C: equivalence of ΔG° (kcal mol⁻¹) by equilibrium (-6.68 ± 0.89) and kinetics (-6.84 ± 0.18 kcal mol⁻¹); equivalence of m° (kcal mol⁻¹ M⁻¹) by equilibrium (1.77 ± 0.24) and kinetics (1.81 ± 0.03). At 20°C: equivalence of ΔG° (kcal mol⁻¹) by equilibrium (-6.80 ± 0.77) and kinetics (-6.56 ± 0.22); equivalence of m° (kcal mol⁻¹ M⁻¹) by equilibrium (1.83 ± 0.13) and kinetics (1.76 ± 0.04). Other parameters are: at 10°C: $RT \ln k_f^{H_2O} = 3.98 \pm 0.10$ kcal mol⁻¹, $m_f = 1.36 \pm 0.03$ kcal mol⁻¹ M⁻¹, At 20°C: $RT \ln k_f^{H_2O} = 4.36 \pm 0.09$ kcal mol⁻¹, $m_f = 1.32 \pm 0.02$ kcal mol⁻¹ M⁻¹.

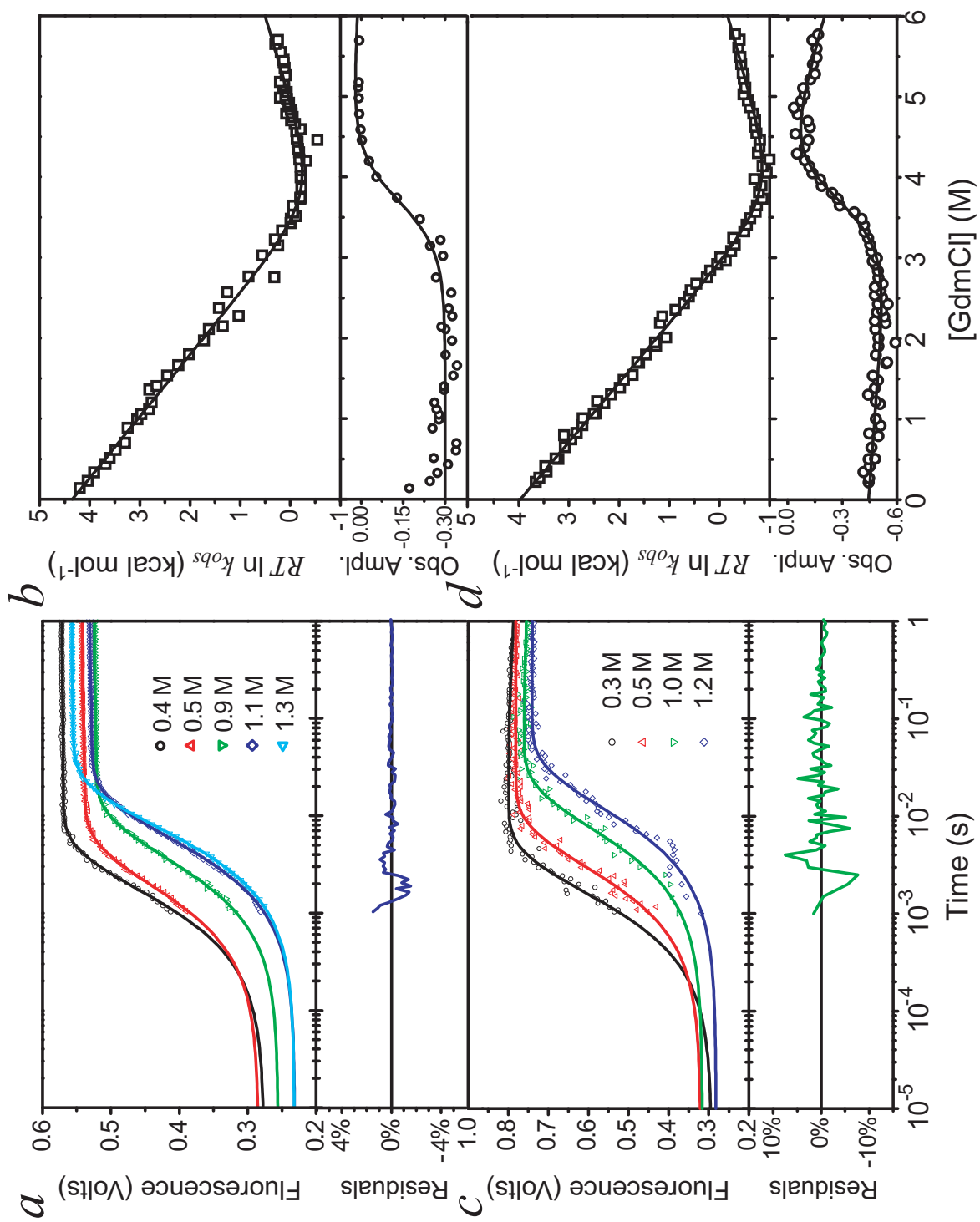
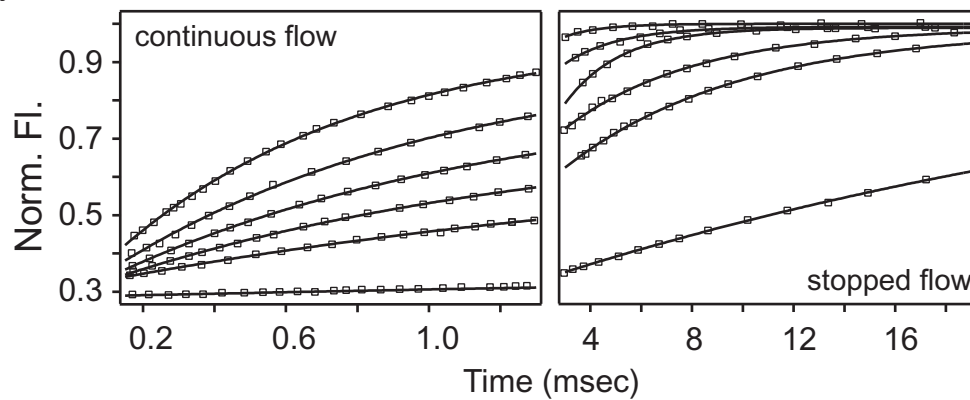
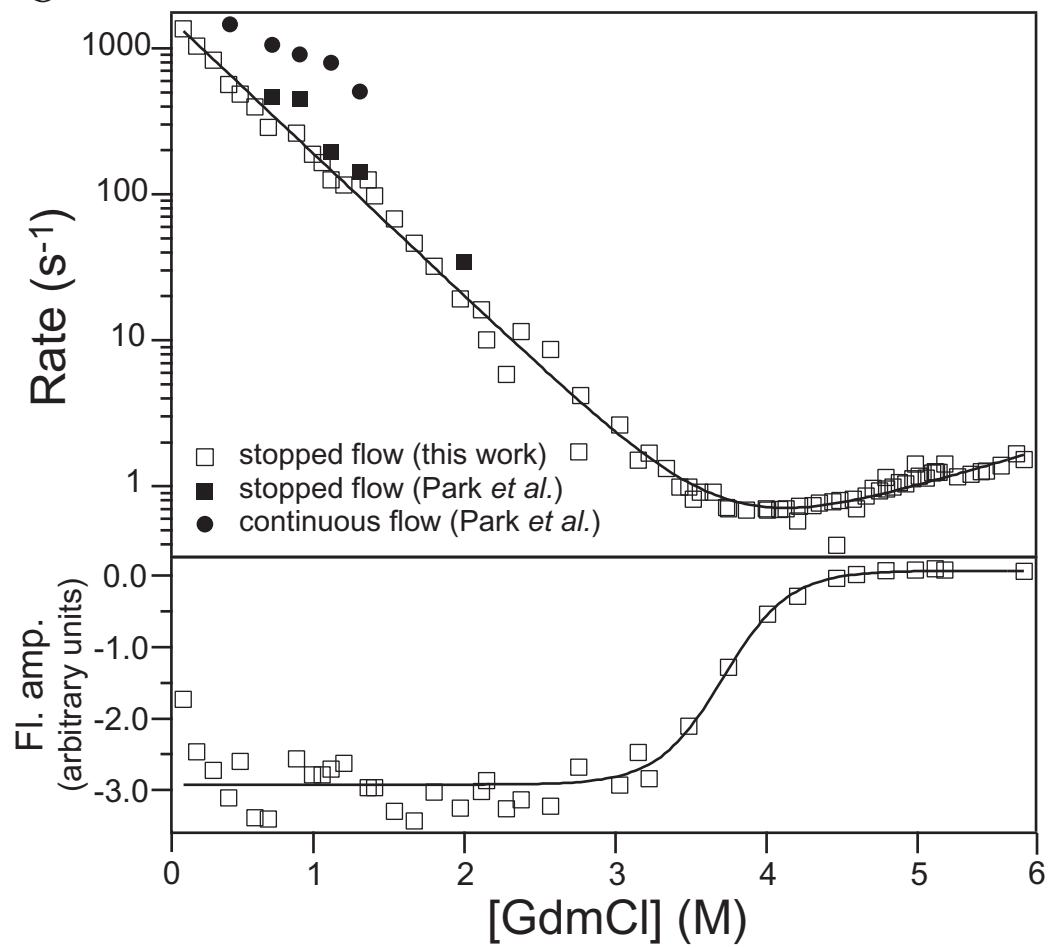


Figure 2.11 Direct comparison of Protein G results to Park et al. (a) Kinetic traces of Park *et al.* are refit to a mono-exponential equation showing that the previous double exponential fits are unjustified. *(b)* Rate constants from the mono-exponential fits are plotted with the 20 °C chevron data from Fig. 2.10 (\square), showing that the continuous flow rates (\bullet) are in error perhaps due to incomplete mixing at the slow flow speeds required to capture a time window suitable for Protein G refolding. Stopped-flow data from Park *et al.* (\blacksquare), however, agree with my stopped-flow data set. Amplitudes are plotted underneath (\square). Single exponential fits to all the data across a broad range of denaturant concentrations yield an ideal, two-state equilibrium curve with no burst-phase.

a*b*

the overall trace appeared to diverge slightly from a single exponential fit. It could be fit better with two exponentials with rates differing by 4-fold. This result was taken to indicate a sequential three-state model (Eq. 2.1) with two barrier crossing events and therefore a distinct populated intermediate. Baker and coworkers (McCallister *et al.*, 2000), however, observed that the folding reactions of GB1 and many mutants are well described by a single exponential without missing amplitude down to 0.8 M GdmCl.

To explore these ambiguities, we conducted ProtG refolding experiments in the presence and absence of 0.4 M Na₂SO₄, but using a stopped-flow apparatus with 1 msec dead-time so that the kinetic data were obtained in a single run and spanned the gap where the bi-exponential break found by Park *et al.* was most pronounced (Fig. 2.11). We find (Fig. 2.10) an accurate mono-exponential fit to the kinetic data with no missing burst-phase amplitude, and linear chevron behavior with no roll-over, down to 0.2 M GdmCl at both 10°C and 20°C. Most definitively, ProtG exhibits rigorous adherence to the two-state chevron criteria; the ΔG parameters determined by equilibrium melting and from the ratio of kinetically measured folding and unfolding rates are in excellent agreement, as are the analogous equilibrium and kinetic *m*-values. This is found in independent data sets at 10°C and 20°C (listed below Fig. 2.10).

2.3.5 IM7

Folding of the four-helix bacterial immunity (Im) proteins, Im7 and Im9, was previously studied (Capaldi *et al.*, 2001; Ferguson *et al.*, 1999; Gorski *et al.*, 2001) using fast continuous-flow and standard stopped-flow methods to probe fluorescence changes

due to a single tryptophan residue. The behavior observed varies with pH and concentrations of Na_2SO_4 , and demonstrates the accumulation of an intermediate.

However, we observed the folding behavior is consistent with a two-state model in the absence of sulfate, e.g. chevron arms are linear with denaturant concentration. Although the hyperchromic burst-phase fluorescence signal, the signature of a sub-millisecond species, is present at the $\sim 20\ \mu\text{M}$ protein concentrations used in the previous studies, the burst signal is greatly reduced at $0.8\ \mu\text{M}$, and is absent at $80\ \text{nM}$. Notice in Fig. 2.12 that simple mono-exponential fits are best applied to the lower concentrations of protein. Moreover, even with a $1\ \text{ms}$ dead-time, evidence of the hyperchromic phase is detected in the kinetic traces at high protein concentrations as shown by the large systematic deviation in the residues when fit with a mono-exponential function. Such systematic deviations are not, however, explained by the signal-to-noise ratio for the measurement, since the deviations occur at high protein concentrations which have 300-fold better signal-to-noise per shot. To obtain equivalent signal-to-noise shots were averaged in approximately a 300:1 ratio.

Additionally, all the H-bond formation, free energy and surface burial can be accounted for in the observed reaction. The equilibrium isotope effect, stability and m -value are recapitulated from the kinetic measurements (Fig. 2.14). The folding kinetic isotope effect remains constant, which is inconsistent with the accumulation of a highly helical intermediate at lower denaturant concentration (see Ch. 3). In the absence of transient aggregation, we find that the folding of IM7 is kinetically two-state. The following experiments were performed to confirm that the accumulated hyperchromic species was in fact due to transient protein aggregation.

Figure 2.12 Protein concentration dependence of IM7 refolding in GdmCl. IM7 was refolded from 6 M GdmCl denatured sample by a 12-fold dilution into refolding buffer (50 mM HEPES, pH 7.0, 10 °C) at six protein concentrations that vary over 2.5 log units. Notice that both the amplitude and shape of the kinetic traces (*top*) change over the concentration range tested in a systematic manner. The amplitude is greatest at high protein concentrations. 6 M GdmCl denatured IM7 (●) is plotted as the normalization reference standard. The hyperchromic shift in amplitude can also be characterized by fits to single exponential functions. The traces which fit best to mono-exponential functions, therefore, are lowest in protein concentration as demonstrated by the residuals plotted (*bottom*) which compare 25 μ M (.....) and 80 nM (—) IM7 refolding data.

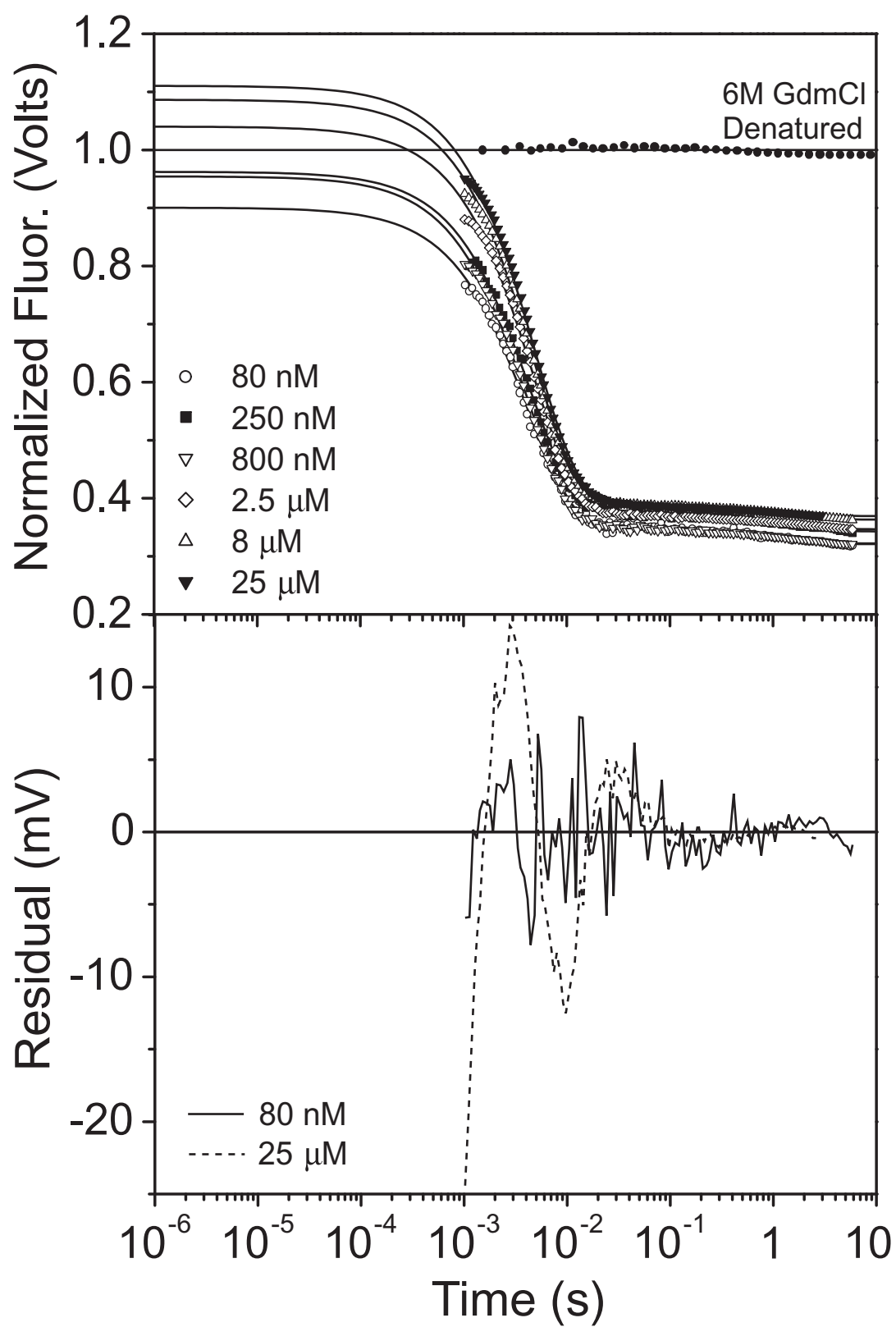


Figure 2.13 Protein concentration dependence of IM7's burst-phase in urea. IM7 was refolded from 6 M urea denatured sample by a 6-fold dilution into refolding buffer (50 mM HEPES, pH 7.0, 10 °C) over 3.5 log units of protein concentration. Normalized amplitudes were estimated by normalizing fluorescence traces to the final folded fluorescence value. Buffer background fluorescence values were subtracted prior to fluorescence rescaling. Accurate determination of the buffer value was determined by linear fits to final fluorescence for five or more protein concentrations. The intercept to such a fit is the background fluorescence of the buffer in the absence of IM7. These data are then fit to a simple binding model where the K_d is 240 ± 50 nM, and the amplitude is 1.04 ± 0.06 .

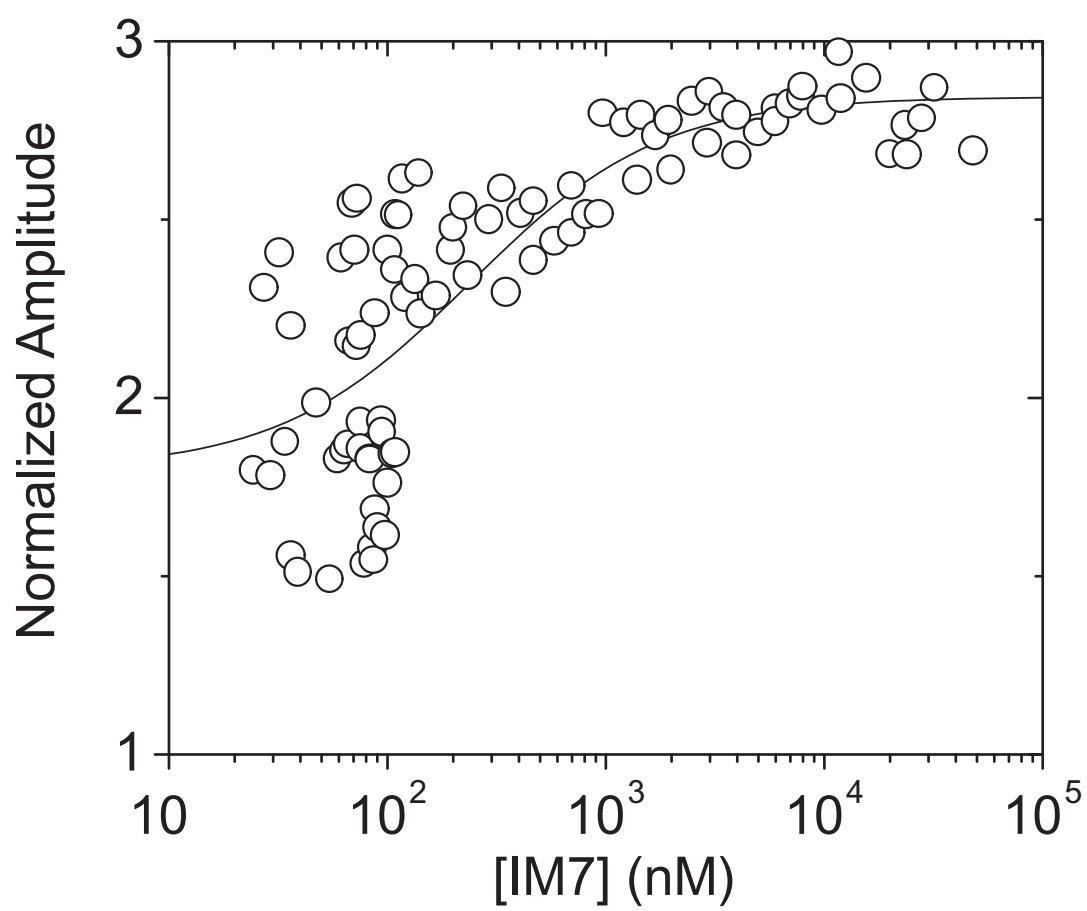
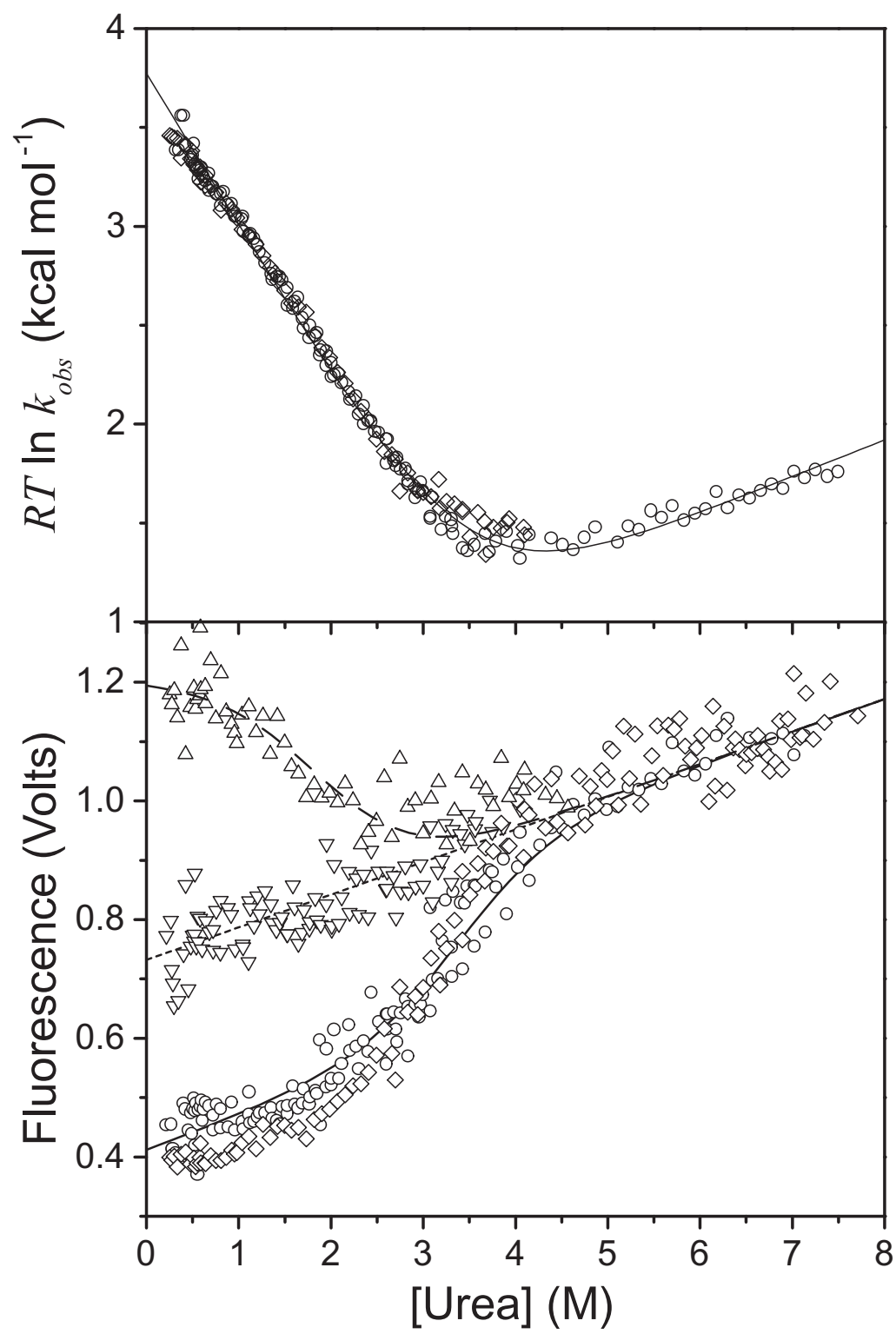


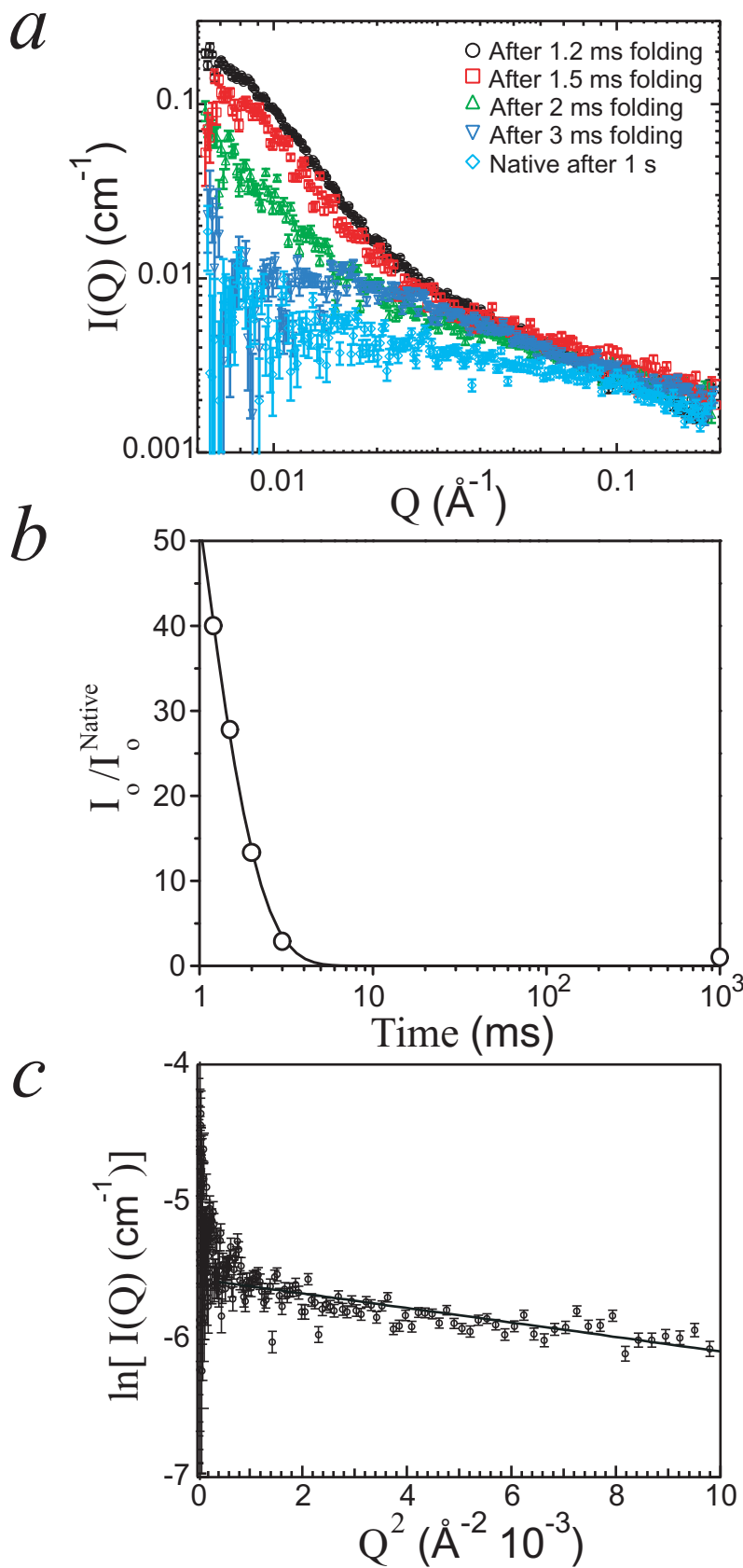
Figure 2.14 The denaturant dependence of IM7 refolding on rates and amplitudes. Urea denaturant dependencies of folding rates and amplitudes were estimated for IM7 under refolding and unfolding conditions in 50 mM HEPES, pH 7.0 at 10 °C. The chevron plots of the folding activation energies show little deviation from near perfect chevron forms (*top*), where two protein concentrations, 20 μM (\diamond) and 160 nM (\circ), were studied. Chevron fit parameters are $\Delta G_f^\ddagger = 3.78 \pm 0.01 \text{ kcal mol}^{-1}$, $\Delta G_{eq} = 3.34 \pm 0.06 \text{ kcal mol}^{-1}$, $m_f = 0.77 \pm 0.01 \text{ kcal mol}^{-1} \text{ M}^{-1}$, and $m^\circ = 0.96 \pm 0.01 \text{ kcal mol}^{-1} \text{ M}^{-1}$. The normalized fluorescence amplitudes, however, (*bottom*) show dramatic differences for the two protein concentrations tested. Final fluorescence values for 20 μM (\diamond) and 160 nM (\circ) protein as well as initial fluorescence values for 20 μM (\triangle) and 160 nM (∇) protein are provided. Notice the hyperchromic shift in initial the fluorescence signal occurs solely at high protein concentration where the change in denaturant concentration appears to melt the putative aggregate.



First, IM7 was refolded by dilution from 6 M urea to 1 M urea across 3.5 orders of magnitude in protein concentration. If the folding process is independent of protein concentration, then the observed hyperchromic burst-phase amplitude should be invariant. However, the observed amplitude is dependent on protein concentration and fits to a simple binding model with a K_d of 240 ± 50 nM. The hyperchromic phase is thus attributable to aggregation, albeit indirectly from fluorescence measurements, where the estimated increase in fluorescence is approximately 56% more in the aggregate than in unfolded monomeric protein. A ~50% increase in fluorescence is typical for the burial of a tryptophan residue in a hydrophobic environment. Therefore, the transient aggregate contains significant burial of hydrophobic surface surrounding the tryptophan chromophore. The fluorescence of the native protein is reduced as compared to the unfolded monomeric protein despite the fact that the tryptophan is buried, since a nearby histidine residue quenches fluorescence in the native fold.

Nevertheless, to better confirm whether the observed hyperchromic burst-phase is due to protein aggregation, we studied the burst-phase species using small angle x-ray scattering (SAXS) in the presence of 0.4 M sodium sulfate. Using the high intensity x-rays generated at Argonne National Lab's Advanced Photon Source (APS), we interfaced the Biologic SFM-400 stopped-flow using a 1 mm quartz capillary as the flow cell where the scattered x-rays were detected on a CCD imaging plate. Since accurate determination of the scattering profiles required ~1 second of data, each time point for the folding reaction was determined using continuous flow mixing. Dead volume, or the volume between the point of rapid mixing and the point of illumination with the focused x-ray beam, when multiplied by the flow rate determines the time point of the folding

Figure 2.15 Continuous-flow SAXS on IM7. a, The scattering profiles for 6 M denatured IM7 sample under continuous-flow after a 6-fold dilution into refolding buffer (50 mM HEPES, 0.4 M Na₂SO₄, pH 7.0, 5 °C). Exposure to x-rays was for 600 to 900 ms. Because the dead volume of the SFM-400 is 12 µL, then flow speeds of 10, 8, 6, and 4 ml s⁻¹ produced the times given in the figure legend. Native IM7 (after 1 second of refolding) was collected in stopped-flow mode where the scattering profile was measured after 500 ms of refolding. The protein concentration (230 µM) was identical for all experiments and buffer scattering profiles were individually subtracted from each profile. Two to six profiles were averaged to produce the data presented. Notice the dramatic increase in intensity at low angle, which is indicative of protein aggregation, for the earliest times. At later times, the intensity at low Q is reduced. *b,* The intensities at Q values of 0.007 were used as a surrogate estimate of relative intensity at zero angle, or I_o / I_o^{Native} . These values were plotted versus time and fit to a mono-exponential function. Notice after 1.2 ms of refolding the protein forms large molecular weight aggregates, which are composed of ~40 monomers. The fit's amplitude is 210±20 with a rate constant of 1360±30 s⁻¹. Finally, Guinier analysis (*c*) of native IM7 after 1 second of refolding was performed. The experimentally derived R_g of 12.6±0.2 Å is within experimental error of the estimated value from the protein's crystal structure. Here I_o fit to $0.00385 \pm 3 \times 10^{-5}$, and $R_g \times Q_{max}$ is 1.28.



reaction. Dye mixing studies had previously characterized that the dead volume is approximately 12 μL .

The scattering profiles at the fastest mixing time available, 1.2 ms, shows a 40 fold increase in the intensity at the lowest Q -value measured as compared to protein that was folded for 1 second, which is indicative of large scale protein aggregation. Moreover, Guinier fits to the 1 second folded scattering profile demonstrates that after the protein has completely folded its measured radius of gyration (R_g) of $12.6 \pm 0.2 \text{ \AA}$ agrees with the value estimated from the proteins crystal structure (12.5 \AA). Furthermore, the aggregation indicated by the increase in intensity at low Q suggests, at first approximation, that the aggregate is composed of at least 40 monomers. The intensity at low Q decreases as the folding reaction progresses in an exponential fashion, demonstrating that the aggregate is disassembled as the individual monomers fold to the native state. Finally, the agreement of the calculated and experimentally derived values for the radius of gyration following one second of folding demonstrates that native IM7 monomers do not interact and form multimeric aggregates.

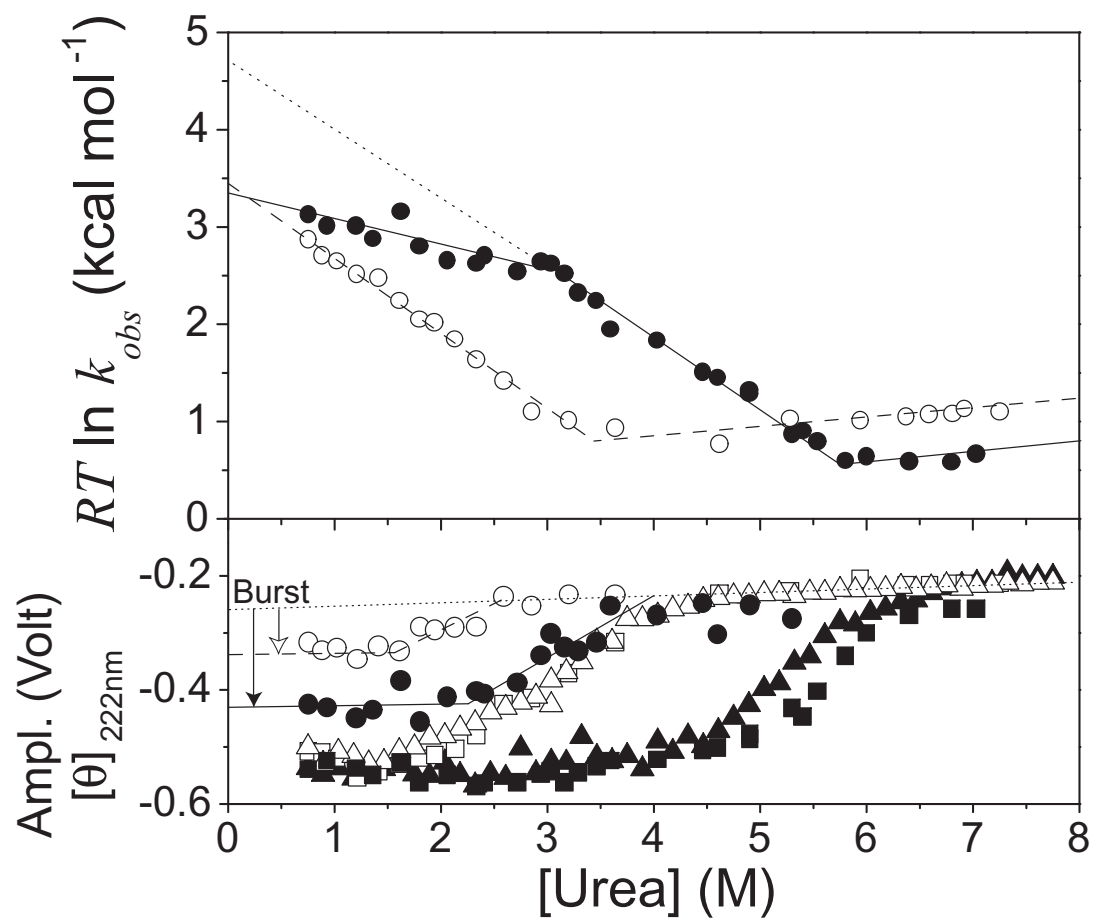
To determine the degree to which secondary structure formed in the aggregate, stopped-flow circular dichroism (CD) experiments were performed. Since experiments were performed where the protein concentration was sufficient to aggregate, the burst-phase CD data reveal that the aggregates possess from 30 to 60% of the native helical content, depending upon the solvent conditions. When refolding in the presence of 0.4 M sulfate, the CD burst-phase is greatest with about 60% helical content.

Previously, the addition of sulfate has been interpreted as a stabilizing additive for folding intermediates. However, since sulfate further drives transient protein aggregation

in IM7's folding pathway and induces significant helical structure, then hydrogen exchange pulse labeling studies conducted in the presence of the additive may contain such transient aggregation. While the effects reported here for IM7 may be specific, it is unclear whether the possibility of transient protein aggregation occurs in HX pulse labeling and HX competition studies conducted on other proteins (Lorch *et al.*, 1999; Lorch *et al.*, 2000; Parker *et al.*, 1997; Sauder & Roder, 1998; Bycroft *et al.*, 1990; Matouschek *et al.*, 1992). The prospect of such aggregation occurring in other systems is likely, however, since NMR experiments require high concentrations of protein, which is known to exacerbate such aggregation as demonstrated for IM7.

Previous work on IM7 revealed compelling data regarding the accumulation of a transiently populated intermediate, which possessed hyperchromicity in fluorescence amplitude (Capaldi *et al.*, 2001; Ferguson *et al.*, 1999; Gorski *et al.*, 2001). However, further characterization of this fluorescence behavior over 3.5 orders of protein concentration demonstrates that the species is in fact a transient aggregate composed of at least 40 monomers after 1 ms of refolding. When folded at low protein concentrations, i.e. 100 nM, the hyperchromic phase vanishes as shown with fluorescence measurements, and IM7's folding pathway is best described by a simple two-state model. Stopped-flow CD experiments have further revealed that large scale transient aggregation may occur with significant helical structure. Thus, future work, such as parallel SAXS studies, on other protein systems may address whether aggregation occurs in the presence of sulfate for other systems in which a putative folding intermediate is proposed to be stabilized by the additive.

Figure 2.16 Stopped-flow CD studies on IM7 refolding. Stopped-flow circular dichroism (CD) experiments were implemented by interfacing the SFM-400 to a Jasco J-715 spectropolarimeter by a custom mount. The 1 mm quartz capillary typically used for SAXS experiments was used as the flow cell. Control experiments confirmed that the observed CD signal at 222 nm (5 nm bandwidth) in the flow cell was consistent with the signal observed in a standard 1 mm pathlength quartz cuvette. Refolding experiments were initiated by a 8-fold dilution of 5.5 M urea denatured IM7 into refolding buffer (50 mM sodium phosphate, pH 7.0, 10 °C) with varying concentration of urea. The observed rates (*top*) as well as the initial and final amplitudes (*bottom*) were recorded both in the presence (*filled symbols*) and absence (*open symbols*) of 0.4 M Na₂SO₄ (final concentration). The final protein concentration was 32.6 μM in all experiments. While the activation energies for folding (*top*) in the absence of 0.4 M Na₂SO₄ (○) showed ideal chevron kinetics, those plotted in its presence (●) were non-ideal and rolled over at 3 M urea. The final equilibrium amplitudes (*bottom*) for IM7 were recapitulated by kinetics and equilibrium experiments in the presence (■ and ▲, respectively) and absence of 0.4 M Na₂SO₄ (□ and △, respectively). The initial CD amplitude prior to refolding without (○) and with (●) Na₂SO₄ shows significant “burst” behavior where the burst is marked by an open-head arrow and filled-head arrow for experiments conducted with and without Na₂SO₄, respectively. Protein aggregates stabilized by Na₂SO₄ are thus more helical in its presence (60% burst) than in its absence (30% burst). Lines provided in the figure are meant to guide the eye.



2.3.6 CD2.D1

Clarke, Parker and coworkers studied the folding of the β -protein CD2.D1 following denaturant dilution into 0.4 M Na₂SO₄ and found deviations from ideal two-state folding behavior, including a kinetic burst-phase, chevron roll-over, and low level HX protection in the burst-phase distributed throughout the measurable β -strands. HX protection factors were less than 6 for 37 of the 41 NH's measured. These observations were interpreted in terms of a burst-phase intermediate with widespread hydrogen bonding (Lorch *et al.*, 1999; Lorch *et al.*, 2000; Parker *et al.*, 1997).

As noted before, these same observations can stem from a burst-phase readjustment of the U state, especially in concentrated sulfate. More definitively, well determined mutational phi values for six core mutants were 0.0 ± 0.1 (Lorch *et al.*, 1999), suggesting that very little side-chain organization has formed in the burst-phase (Lorch *et al.*, 2000). Further, native-state HX measurements found no stable hydrogen bonded intermediate (Parker & Marqusee, 2001). These results are consistent with the view that, as for other proteins, the burst-phase signals register a solvent-dependent response of the denatured state rather than the formation of a distinct intermediate.

2.3.7 Summary of early intermediates

The inferred existence of stable burst-phase folding intermediates has depended on indirect evidence, especially burst-phase F1 signals and chevron roll-over. These and other indications have alternative explanations. The only structural information available, widespread minimal HX protection in concentrated Na₂SO₄, appears to result from chain

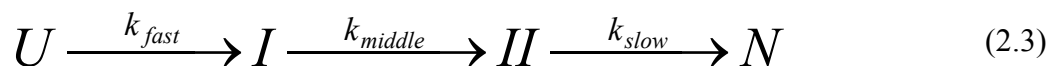
condensation (U to U') rather than regular H-bonded secondary structure in stable intermediate forms. U' may be composed of transient protein aggregation as observed for IM7; therefore, future experiments should study folding at low protein concentrations (100 nM) and rule out transient aggregation with parallel SAXS studies.

2.4 Late Intermediates

The initial barrier hypothesis suggests that late intermediates accumulate only when a misfolding-dependent barrier is inserted, especially so in small proteins. Some apparent exceptions are considered here.

2.4.1 Cytochrome *c*

In addition to the burst-phase intermediate (I in Eq. 2.3), Akiyama et al (Akiyama *et al.*, 2000) proposed an obligatory late intermediate (II in Eq. 2.3) in Cyt *c* folding, suggested by a minor slow CD phase



Other results point to a different interpretation involving folding heterogeneity, due to an optionally inserted, pH dependent misfold-reorganization barrier.

In earlier work, multiple independent kinetic methods consistently showed that Cyt *c* can fold all the way to N in a two-state manner, with rate constant close to k_{middle} in Eq. 2.3 (temperature and denaturant corrected) (Sosnick *et al.*, 1994). Cyt *c* can also be made to fold in a three-state manner. In this case, an intermediate composed of the N- and

C-terminal helices accumulates at the rate of k_{middle} in Eq. 2.3, and then more slowly folds to N (Sosnick *et al.*, 1996a; Sosnick *et al.*, 1994). A number of observations indicate that this can be understood in terms of the insertion of an optional error-repair barrier in the folding sequence. The barrier insertion is pH dependent (histidine-related). At any given pH some fraction of the protein population folds to N in a rapid two-state manner (k_{middle}) while another fraction experiences the inserted barrier and reaches N more slowly (k_{slow}), in a three-state manner. Removing the two peripheral histidines of Cyt *c* removes the slow folding barrier at any pH (Fig. 2.3g). Folding past the inserted barrier is promoted rather than slowed by denaturant. Similarly, Rousseau and coworkers used rapid mixing resonance Raman experiments to demonstrate that the folding of Cyt *c* all the way to N occurs at a denaturant corrected rate of $k_{middle} = \sim 2,000 \text{ s}^{-1}$ at pH 5, 22°C (Takahashi *et al.*, 1997; Yeh & Rousseau, 1998; Yeh *et al.*, 1997), unless a transient pH-dependent mis-ligated state develops before or during folding. The different CD phases seen by Akiyama *et al.* have approximately these same rates and pH dependence. There is some discrepancy among the various slow rates measured; protein aggregation or proline mis-isomerization potentially plays a role.

All of these results support the conclusion that the middle phase observed by Akiyama *et al.* (Eq. 2.3) represents two-state folding to N for the major population fraction while the slowest kinetic phase reflects an optionally inserted misfold-reorganization barrier rather than an obligatory late barrier. To further test this view, we folded Cyt *c* under the same conditions used by Akiyama *et al.*, that is, jumping from U at pH 2 to pH 4.5, measuring initial chain collapse by FI quenching (Trp59 to heme

proximity) and final native state acquisition by 695 nm absorbance (Met80-S to heme Fe ligation; Figs. 2.3*h,i*). When the burst-phase sensed only by Fl quenching is subtracted, the probes for initial collapse (Fl) and final folding (A_{695}) exhibit identical kinetics at 10 and 22°C, demonstrating two-state folding for the large majority of the refolding population. A small fraction either mis-ligates heme or aggregates, reaching N more slowly.

In summary, observations on fast and slow Cyt *c* folding have been extensive and complex. The overall view that emerges from the present considerations is as follows. In general, three kinetic phases can be seen. When folding is initiated by denaturant dilution, a sub-millisecond burst-phase (~50 μ sec) occurs that represents a solvent-dependent readjustment of the U state to U'. The N state can be reached in ~1 msec for a large portion of the population, but the fraction depends on initial and final pH. The remaining fraction is blocked and folds to N more slowly. When folding is initiated in the absence of denaturant (e.g. pH jump, CO photolysis, iron reduction), again a burst-phase dependent on chain contraction and fast change in heme ligation can produce various optical signals and low level HX blocking due to sodium sulfate, but no helical CD. Subsequent native state acquisition occurs at two different rates in populations that are either blocked or not, depending on initial and final pH which determines heme mis-ligation and misfolding. Additional confounding variables include aggregation, His-heme mis-ligation, which can cause chevron roll-over in early folding and possibly also later folding artifacts, as can proline mis-isomerization and the fact that pre-bound heme ligands take some time to escape from the fully folded protein.

2.4.2 *Barstar*

As a proxy for many other proteins, barstar is known to fold via intermediates that accumulate due to barriers inserted by non-native proline isomers (Agashe *et al.*, 1995; Bhuyan & Udgaonkar, 1999). This leads to heterogeneous folding in which a sub-population with native isomers folds directly to the native state, demonstrating that the late barrier that causes intermediate accumulation does not represent an intrinsic folding event but rather a probabilistic misfold-reorganization step.

2.4.3 *Summary of later intermediates*

Late on-pathway intermediates accumulate in Cyt *c* and many other proteins due to histidine mis-ligation or proline mis-isomerization. These intermediates occur after the initial intrinsically rate-limiting step. The later barrier that blocks folding and causes intermediate accumulation reflects an optionally inserted error-correcting reorganization process rather than an intrinsic folding step.

2.5 Discussion

Current issues in protein folding concern the nature of rate-limiting barriers: Are barriers early or late? Is their character intrinsic or optional? How large are the barriers, that is, how fast does folding go? Are folding intermediates, created by barriers, discrete or continuous? Or are intermediates obligatory or optional? Or are they constructive or obstructive? The Initial Barrier Hypothesis—that the folding of globular proteins is intrinsically rate-limited by the first committed on-pathway step (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995)—touches on all of these issues.

In the initial barrier view, proteins should tend to fold in an apparent two-state manner, without the accumulation of stable intermediates. Stable intermediates cannot occur before the initial obligatory barrier; later intermediate accumulation due to stable misfolding errors is minimized by the extensive organization of the main chain in the initial transition state nucleus. This expectation seems most compelling for small proteins for which the initial nucleus must represent a large part of the chain so that the folding problem is essentially solved in the *initial* rate-limiting step. Here we have analyzed existing data and described new results for a number of small proteins that seemed to violate this principle by populating stable intermediates before the initial barrier or by encountering obligatory large folding barriers after it. Closer examination finds that small proteins almost universally behave in accord with the hypothesis.

With respect to the structure of rate-limiting transition states, early theoretical (Abkevich *et al.*, 1994; Dill *et al.*, 1993; Moult & Unger, 1991; Thirumalai & Guo, 1995) and experimental (Itzhaki *et al.*, 1995a; Itzhaki *et al.*, 1995b; Matheson & Scheraga, 1978) results pointed to distinct folding nuclei involving a small number of important side chain interactions. Our view has been that the folding transition state nucleus is extensive and topologically native-like; the conformational search for such a structure occurs as the very first on-pathway step; and this time-consuming search ultimately limits the rate at which proteins can fold. Evidence for these views came initially from studies of rate-limiting steps in the folding of Cyt *c* (Englander *et al.*, 1998; Sosnick & Englander, 1996; Sosnick *et al.*, 1996b; Sosnick *et al.*, 1994; Sosnick *et al.*, 1995). The more recent literature confirms the extensive nature of the rate-limiting nucleation process (Jackson, 1998) and points to native-like topology in the folding transition state (Chiti *et al.*, 1999; Goldenberg, 1999; Martinez *et al.*, 1999; Plaxco *et al.*, 1998; Riddle *et al.*, 1999).

2.5.1 *Stable intermediates*

It is often stated that intermediates do not accumulate in folding, because they are not stable relative to U, or because they do not exist. In fact, many cases of stable partially folded intermediates are now known. Direct structural evidence for such forms has come especially from hydrogen exchange investigations of folding intermediates in kinetic and equilibrium modes (Englander, 2000; Rumbley *et al.*, 2001). Also stable analogs of partially folded proteins have been prepared by synthesis (Chamberlain *et al.*, 1999). These observations show that stable intermediates, which turn out to represent partial replicas of the native protein at greater or lesser degree of advancement, occur ubiquitously, even when kinetic folding and equilibrium unfolding appear two-state. It thus seems unlikely that the prevalence of two-state folding can be explained by the view that stable intermediates do not exist. The Initial Barrier Hypothesis considered here provides an alternative reason for two-state folding. Intermediates do not accumulate in kinetic folding, because they form only after the initial rate-limiting step (Rumbley *et al.*, 2001).

The absence of stable intermediates on the unfolded side of the initial folding barrier appears to be due to thermodynamic constraints. Collapse to a distinct species is inhibited by the loss of conformational entropy and the small number of possible stable intermediates. Enthalpically, the burial and desolvation of polar groups, both main chain and side chain, can only occur in a context that satisfies their hydrogen bonding requirements, with each failure costing several kcal mol⁻¹. These exacting requirements make the search for any condensed structure uphill in free energy from U. Also they bias chain condensation toward a topologically native-like organization defined by native-like structure-determining hydrogen bonds (Krantz *et al.*, 2000; Krantz *et al.*, 2002; see Ch. 3).

These same theoretical constraints and their bias toward native-like topology apply equally to the formation of the rate-limiting transition state nucleus, in agreement with experimental observations. A native-like character for the nucleating transition state is also required by its functional role. It must organize multiple interactions that can guide and stabilize the subsequent construction of native-like intermediates in a downhill manner. It is this uphill to downhill switch that defines a “correctly” collapsed, i.e. native-like, conformation and makes it the effective transition state.

2.5.2 Late barriers

Late rate-limiting barriers can intrude and cause normally occurring but corrupted intermediates to accumulate. These barriers most often represent misfolding errors that require some time-consuming reorganization process, and such late barriers do not reflect intrinsic barriers to normal folding. Known examples include *cis/trans*-proline and non-proline mis-isomerization errors common to many proteins, such as pH-dependent histidine-heme mis-ligation in Cyt *c*. The adventitious nature of later barriers is abundantly demonstrated by the fact that some population fraction is almost always seen to fold directly to N in a two-state manner, even when the folding of another population fraction is three-state, giving rise to the widely observed heterogeneity of protein folding.

2.6 Conclusion

The present analysis shows that the folding of small globular proteins appears almost universally to be kinetically two-state, in spite of the prevalence of on-pathway intermediates. The initial barrier hypothesis leads one to expect this otherwise surprising

result (Sosnick *et al.*, 1996b). The hypothesis proposes that proteins enter the folding pathway only when the search through U-state space by chance finds a topologically native-like conformation that can serve as a transition state able to nucleate downhill native-like structure formation. Stable native-like intermediates essentially do not form before the initial large scale search step succeeds, and they do not accumulate afterwards, because later intrinsic barriers represent smaller scale conformational searches that are not rate-limiting. Exceptions reflect accidental or imposed misfolding events. This discussion is limited to small proteins for which these considerations seem most compelling. One can expect that much the same principles should apply also to larger proteins where, however, the probability of misfolding and late intermediate accumulation is likely to be greater.

2.7 Methods

2.7.1 Proteins

Horse heart Cyt *c* (Akiyama *et al.*, 2000) of the highest grade was purchased from Sigma Chemical Company. The CNBr fragments were described before (Sosnick *et al.*, 1997). The preparation of the H26N/H33N variant of Cyt *c* will be described elsewhere. Ub was prepared as described (Krantz & Sosnick, 2000). The 6-His B1 domain of protein G was prepared from standard *E. coli* expression lysates using an expression vector kindly provided by D. Baker. For IM7, a plasmid construct used pRSET B (Invitrogen) with total DNA synthesis (Operon technologies) for the published protein sequence (Ferguson *et al.*, 1999). See App. 2 for further details on protein sequences and procedures.

Purification of these constructs from standard *E. coli* expression lysates used a combination of anion or cation exchange chromatography, nickel affinity chromatography, and C18 reverse phase HPLC. All proteins were analyzed by mass spectrometry.

Deuterated versions of Cyt *c* were made by incubating the protonated protein in deuterated 7.4 M GdmCl buffer. On dilution into H₂O, amides remain deuterated for tens of seconds. Acid denatured Cyt *c* was prepared by extensive dialysis against 10 mM hydrochloride. Folding was initiated by a 21:4 dilution with 300 mM sodium acetate, pH 4.8 (final pH 4.5).

2.7.2 *Equilibrium measurements*

Standard free energies of folding, ΔG° , were determined from GdmCl denaturation profiles monitored by circular dichroism spectroscopy at 222 nm using a Jasco J-715 spectropolarimeter with a computer interfaced titrator. Measurements were conducted with 2 nm resolution and a 1 cm path length cuvette. Peptide concentrations were 5 μ M, and experiments were carried out in 20 mM sodium acetate.

2.7.3 *Kinetic measurements*

Rapid mixing experiments used a Biologic SFM4 stopped-flow apparatus as described (Krantz & Sosnick, 2000). To improve the dead-time performance, only the bottom 20% of the 0.8 mm cross-section FC-08 cuvette was illuminated through a fiber optic cable connected to a PTI A101 arc lamp. The dead-time was calibrated using a dye

quenching reaction. Fluorescence spectroscopy used excitation and emission wavelengths of 280-290 nm and 300-400 nm, respectively.

For the single-jump measurements, starting from the fully unfolded state (4-7 M GdmCl) or the folded state (1 M GdmCl), folding or refolding reactions were initiated by dilution to yield the desired denaturant concentration. Total volume per shot was 0.25 ml, except for dilutions greater than 1:8 where the total volume delivered was increased up to 0.5 ml. This extra volume was required in order to ensure that the observed folding reaction was from a fresh protein solution that was uncontaminated by mixing with refolding buffer between shots. Failure to do so resulted in a minor, pre-folded population and an accompanying change in the fluorescence signal, as determined by continuous-flow control experiments. For the double-jump measurements, native ubiquitin in 2 M GdmCl was mixed with 7.4 M GdmCl to a final concentration of 4.7 M GdmCl. After a ten second delay (about ten unfolding half-lives), the ~40 μ l denatured protein solution contained in the aging loop was diluted with buffer to yield the final denaturant concentration. Continuous-flow measurements used the same protocol as the single-jump measurements, but the flow-speed was reduced to produce the desired dead-time, $\tau_d = (\text{dead volume})/(\text{flow-speed})$. Ubiquitin protein concentrations were 0.7-2.6 and 1.3 μ M for the single-jump and double-jump measurements, respectively. Two to six time courses were averaged at each reaction condition.

The IM7 burst-phase species was studied kinetically using small angle x-ray scattering (SAXS) in the presence of 0.4 M sodium sulfate. The high intensity x-rays generated at Argonne National Lab's Advanced Photon Source (APS) were utilized by interfacing the Biologic SFM-400 stopped-flow via a custom mount. A 1 mm quartz

capillary flow cell was implemented, where the scattered x-rays were detected on a CCD imaging plate. Since accurate determination of the scattering profiles required ~1 second of data, each time point for the folding reaction was determined using continuous flow mixing. Exposure to x-rays was actually for 600 to 900 msec. Native IM7 (after 1 sec of refolding) was collected in stopped-flow mode where the scattering profile was measured for ~1 sec after 500 ms of refolding. The protein concentration (230 μ M) was identical for all experiments and buffer scattering profiles were individually subtracted from each profile. Two to six profiles were averaged to produce the data presented.

Stopped-flow circular dichroism (CD) experiments were implemented by interfacing the SFM-400 to a Jasco J-715 spectropolarimeter by a custom mount. The 1 mm quartz capillary typically used for SAXS experiments was used as the flow cell. Control experiments confirmed that the observed CD signal at 222 nm (5 nm bandwidth) in the flow cell was consistent with the signal observed in a standard 1 mm pathlength quartz cuvette.

2.7.4 Data analysis

Kinetic data were analyzed using the chevron analysis (Matthews, 1987) with the free energy of equilibrium folding and the activation free energy for kinetic folding and unfolding assumed to be linearly dependent on denaturant according to the standard equations. Parameters were fit using a non-linear least-squares algorithm implemented in the Microcal Origin software package.

3

Protein Hydrogen Bond Network Formation And Backbone Amide Isotope Effects

3.1 Abstract

The following chapter describes the development of procedures that determine when hydrogen bonds form under fast, two-state folding conditions using equilibrium and kinetic deuterium/hydrogen isotope effects in backbone amides. A library of model proteins is analyzed by this method, including λ -repressor and arc repressor variants, the dimeric and cross-linked versions of the GCN4-p1 coiled coil, two versions of cytochrome *c*, colicin immunity protein 7, proteins L and G, mammalian ubiquitin, acylphosphatase, chymotrypsin inhibitor II, and a *src*-SH3 domain. The equilibrium isotope effect is quantified by the “ C_m experiment,” upon which the change in equilibrium stability of a protein is monitored usually by far UV circular dichroism in a continuous fashion during the hydrogen exchange process. Stopped-flow kinetics experiments ascertain not only the kinetic isotope effect of deuterating backbone amides but also the equilibrium isotope effect, which further corroborates the isotope effect. The change in equilibrium free energy accompanied by backbone deuteration is highly correlated to secondary structure composition when H-bonds are divided into two classes: α -helical and β -sheet or long range. The number of helical H-bonds is well correlated with an average equilibrium isotope effect of $8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ H-bond}^{-1}$ where

protium is more stabilizing than deuterium. However, β -sheet and long range H-bonds have little, if any, isotope effect. Global kinetic isotope effects found in helical proteins support our hypothesis that hydrophobic association cannot be divorced from helix formation in the transition state. Extensive backbone desolvation is a critical component of the TS, with helix formation providing an efficient means to satisfy the H-bond requirements of hydrophobically collapsed environment. Therefore, folding models describing an incremental build-up of structure where hydrophobic burial and H-bond formation occur commensurately are more consistent with these data than are models which posit the extensive formation of one quantity before the other.

3.2 Introduction

Biological reactions involving hydrogen (H) often can be dissected using deuterium (D) substitution (Cleland, 1995; Northrop, 1981; Schowen & Schowen, 1982). The doubling of the hydrogen's mass and accompanying change in zero-point vibrational energy, perturbs chemical equilibria and kinetics and enables the elucidation of mechanism and transition state (TS) structure. Such a perturbation is commonly termed an isotope effect.

In protein folding, kinetic amide isotope effects directly assay for the extent of H-bond formation in the TS and overcome some of the limitations of standard mutational ϕ -analysis, where the degree of interaction of a mutated side chain is assayed and specific information about H-bond formation is inferential. However, only a few D/H amide isotope effect studies were conducted under constant bulk solvent conditions

Figure 3.1 Three-dimensional renderings of model proteins. Secondary structures are colored blue for α -helix and green for β -sheet. GCN4 **(a)**, cytochrome *c* **(b)**, Arc repressor **(c)**, λ -repressor **(d)**, IM7 **(e)**, ubiquitin **(f)**, CTACP **(g)**, Protein G **(h)**, CI2 **(i)**, SH3 **(j)**. ProtL (not shown) is topologically identical to ProtG. Renderings were created in the Swiss-Prot Protein Viewer (Glaxo Wellcome Experimental Research) and the Persistence of Vision ray trace program (POV-Ray).

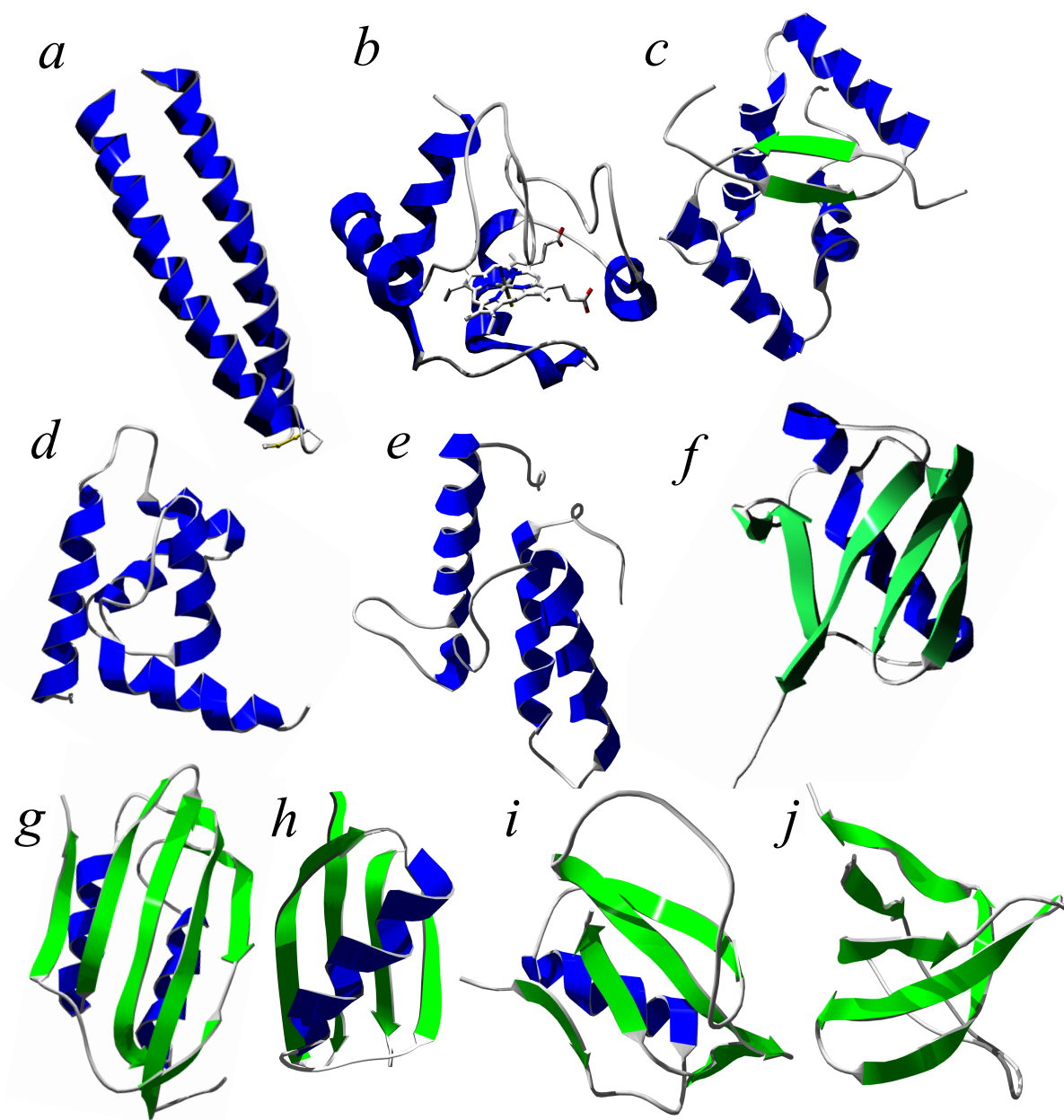


Table 3.1 Model systems for backbone amide isotope effect studies

Protein Description ^a	Species/Source	Abbr.	Class	PDB	Hydrogen Bonds ^b			
					α^d	β	Turn	Other
arc repressor (6-His tag)	Enterobacteria phage P22	Arc	α^e	1ARR	37	7	3	0
arc repressor (6-His tag, R31M, E36Y, R40L)	Enterobacteria phage P22	Arc ^{MYL}	α^e	1MYL	37	7	3	0
colicin immunity protein 7	<i>Escherichia coli</i>	IM7	α	1CEI	39	0	0	12
cytochrome <i>c</i>	<i>Equus caballus</i>	cyt <i>c</i>	α	1HRC	35	0	0	16
cytochrome <i>c</i> (H26N, H33N)	<i>Equus caballus</i>	cyt <i>c</i> ^{pWT}	α	1HRC	35	0	0	16
GCN4-p1' dimeric (D7A, Y17W)	<i>Saccharomyces cerevisiae</i>	GCN4 ^d _{D7A}	α	2ZTA	65	0	0	0
GCN4-p1' (D7A, Y17W) N-terminal cross-link	<i>Saccharomyces cerevisiae</i>	GCN4 ^{Nx} _{D7A}	α	N/A	59	0	0	0
GCN4-p1' (K3H, D7H, Y17W) C-terminal cross-link	<i>Saccharomyces cerevisiae</i>	GCN4 ^{Cx} _{biH3,7}	α	N/A	59	0	0	0
monomeric λ -repressor protein CI ^c	Bacteriophage lambda	λ^{GG}	α	1LMB	51	0	0	5
monomeric λ -repressor protein CI (G46A, G48A) ^c	Bacteriophage lambda	λ^{AA}	α	1LMB	51	0	0	5
chymotrypsin inhibitor II	<i>Hordeum vulgare</i>	C12	α/β	2CI2	11	13	2	7
ubiquitin (F45W)	<i>Homo sapiens</i>	Ub	α/β	1UBI	17	19	6	3
erythrocyte acylphosphatase (common type)	<i>Homo sapiens</i>	CTACP	α/β	2ACY	23	31	3	2
B1 immunoglobulin γ binding domain (6-His tag)	<i>Streptomyces griseus</i>	ProtG	α/β	2GB1	14	18	3	0
B1 immunoglobulin γ binding domain (6-His tag)	<i>Finegoldia magna</i>	ProtL	α/β	2PTL	11	18	3	0
src-SH3 domain	Avian sarcoma virus	SH3	β	1SHF	0	19	5	9

^aSee Apps. 1 and 2 for more detailed information on these proteins.

^bOnly hydrogen bonds that contain a backbone amide are considered.

^cSequence includes only residues 6-85 which make a monomeric version of λ (see Sec. 3.10.1).

^dIncludes a minor population of 3_{10} helical H-bonds.

^eArc has a small β -sheet involving between 4 and 7 H-bonds (depending on the protein structure) but is classified here as α -helical.



(Itzhaki & Evans, 1996; Kentsis & Sosnick, 1998; Krantz *et al.*, 2000; Krantz *et al.*, 2002; Parker & Clarke, 1997). Even though the innumerable non-covalent interactions that comprise a protein should afford an isotope effect (Wade, 1999), such investigations have been largely discouraged by the small isotope effects discovered in small molecule hydrogen bonds (Jancso, 1974), which predicted that the D/H isotope effects in protein hydrogen bonds would be immeasurable due to a cancellation of entropy and enthalpy terms. Furthermore, when a protein folding reaction is conducted in liquid water, hydrogen bonds merely change partners; amides formerly bound to water in the unfolded state subsequently bind to carbonyl in the native (Fig. 3.3). This change of partners was expected to diminish the relative importance of H-bonds to protein stability as compared to other stabilizing forces, notably the hydrophobic effect (Dill, 1990; Kauzmann, 1959; Tanford, 1980), yet the importance of H-bonds in protein stability remains hotly debated (Avbelj *et al.*, 2000; Dill, 1990; Pace *et al.*, 1996; Shi *et al.*, 2002; Yang & Honig, 1995). Recent quantum mechanical estimates of D/H substitution in H-bonds predict isotope effects of 0.1 to 0.2 kcal mol⁻¹ isotope effects (Scheiner, 2000), but this calculation represents only half of the H-bond reaction expected for a protein folding in liquid water as represented in Fig. 3.3.

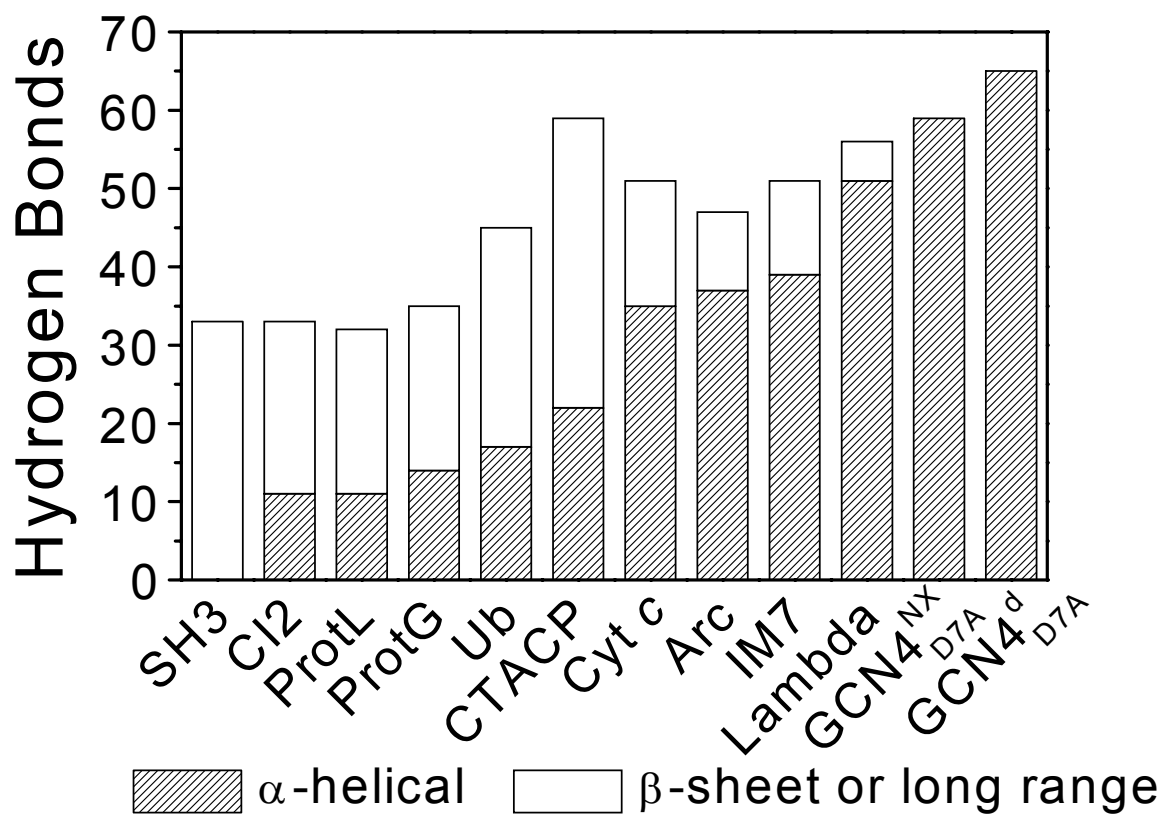
Studies in this chapter are meant to test whether backbone amide isotope effect measurements can elucidate H-bond formation during protein folding and within transition state ensembles. Firstly, can the isotope effect be reproducibly quantified in protein folding equilibrium and kinetics experiments for a simple model system like the purely α -helical GCN4 coiled-coil? Secondly, can isotope effect studies on proteins with more complex, mixed α/β topologies provide reliable global information about H-bond

network formation in both α -helix and β -sheet secondary structures? Thirdly, are H-bond network formation and surface area burial concomitant folding processes that must be sustained in equal proportions in the TS for α -helical proteins? Fourthly, if H-bond network formation and surface area burial are coincident, cooperative processes, what general aspects of folding pathways and protein stability are related to surface burial?

These issues are addressed by analyzing a library of small, two-state folding proteins, which span a wide variety of secondary structure compositions (Fig. 3.2). These model systems (Table 3.1, Fig. 3.1) include three classes: (1) largely α -helical proteins, (2) mixed α/β proteins, and (3) β -sheet proteins. The largely α -helical proteins include: λ -repressor (λ); cytochrome *c* (Cyt *c*); N-terminally cross-linked, C-terminally cross-linked, and dimeric coiled coil (GCN4^{Nx}_{D7A}, GCN4^{Cx}_{biH3,7}, and GCN4^d_{D7A}, respectively); colicin immunity protein 7 (IM7); and dimeric Arc repressor (Arc). The mixed α/β proteins include: chymotrypsin inhibitor II (CI2), common type acylphosphatase (CTACP), ubiquitin (Ub), Protein L (ProtL), and Protein G (ProtG). Finally, the purely β -sheet protein class is represented by a *src*-SH3 domain (SH3).

I will explore the correlation between $\phi_f^{\text{D-H}}$, which is interpreted as the fraction of H-bonds formed in the TS, and the fraction of denaturant sensitive surface area buried in the TS, assessed from m_f/m^0 -values, in GCN4, Cyt *c*, λ -repressor, IM7, and Arc repressor, which are α -helical model systems. These proteins' m_f/m^0 -values span a wide range (0.4 to 0.9). Furthermore, a mutant form of Arc repressor and IM7 are folded under

Figure 3.2 Hydrogen bond composition for small two-state model proteins. Hydrogen bond composition is coarsely divided into helical hydrogen bonds () and β -sheet hydrogen bonds (), which includes a minor population of long range hydrogen bonds. Side chain to backbone H-bonds are ignored, as they are small in number, mostly solvent exposed, and rapidly exchange as compared to backbone amide hydrogen bonds (Bai *et al.*, 1993). Abbreviations for the proteins are given in Table 3.1.



different pH conditions, allowing the proportion of surface area buried in the TS to be shifted as an internal control. In total, these systems sample the correlation of secondary structure formation and tertiary structure burial at values from 40% to 90% surface area burial, which is as wide of a sampling as can be expected for the parameter. Results on these proteins, and the implications for various general folding models are discussed.

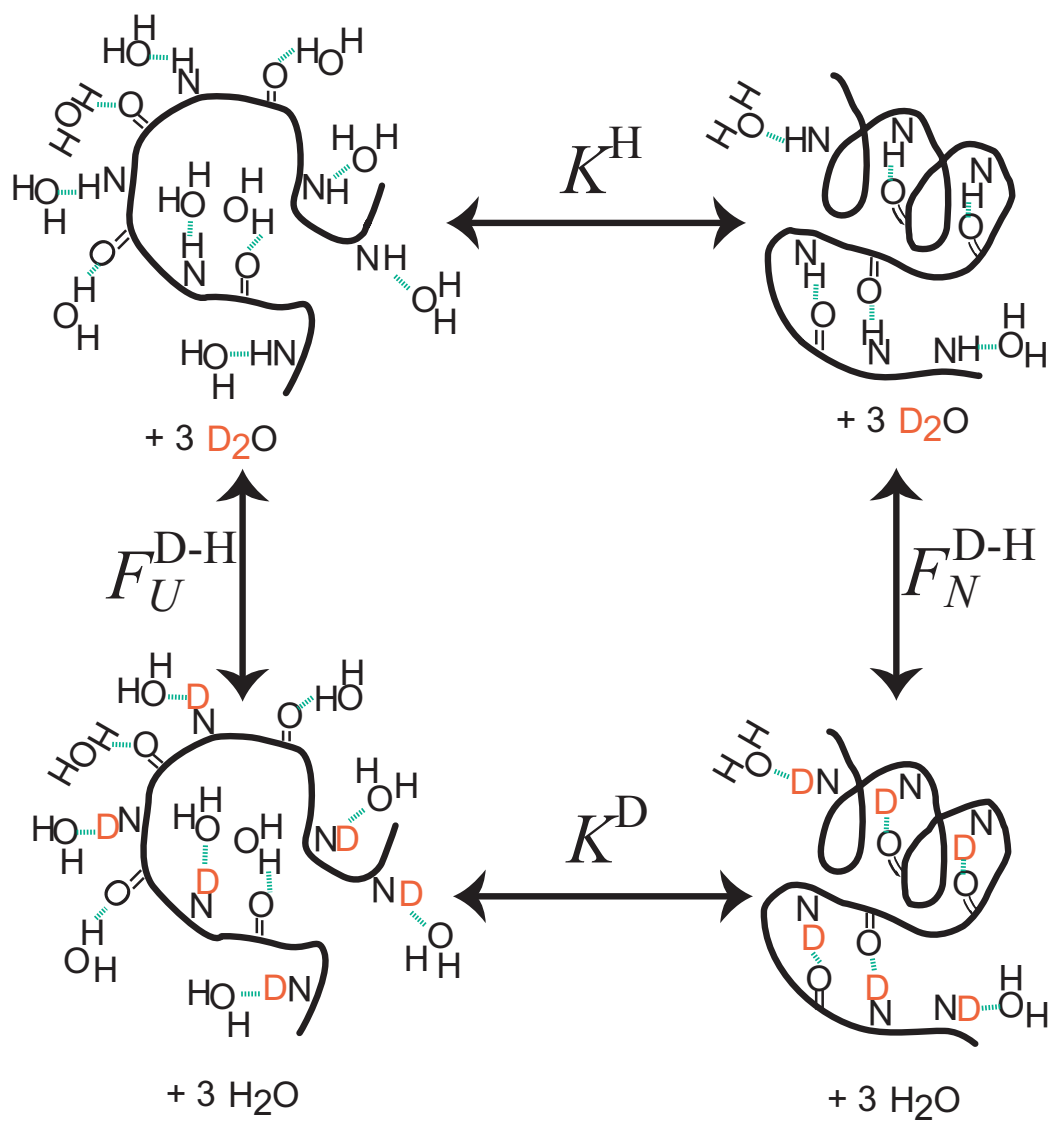
3.3 Equilibrium isotope effects

Protein stability depends upon bulk solvent isotope composition (Itzhaki & Evans, 1996; Krantz *et al.*, 2000; Makhatadze *et al.*, 1995; Parker & Clarke, 1997). Changes in stability may be separated into an isotope effect due to the change in the strength of backbone amide-to-carbonyl and an isotope effect due to a change in the isotopic content of bulk solvent. The bulk solvent isotope effect is presumably attributed to both a change in the hydrophobicity of the water (i.e. the ability of ice-like water-water H-bonds to solvate apolar surface) as well as the strength of water-to-carbonyl H-bonds (Fig. 3.3).

Here the primary interest is in determining when backbone H-bonds form, and hence, only in the change in stability upon amide substitution when measured under constant bulk solvent conditions. This quantity, $\Delta\Delta G_{eq}^{D-H}$, is a result of the differential preference for deuterium in amide-to-water H-bonds in the unfolded state versus native, amide-to-carbonyl H-bonds in the folded state (Fig. 3.3)



Figure 3.3 A thermodynamic cycle for amide isotope effects. The equilibrium isotope effect, $\Delta\Delta G_{eq}^{D-H}$, is the relative stability of protonated protein, K_{eq}^H , compared to the stability of deuterated protein, K_{eq}^D . The folded proteins are on the right side (illustrated as a single α -helix and β -turn) and the unfolded proteins are on the left side (hydrated random coil). All amide and carbonyl sites in the unfolded state are hydrogen bonded to water (green dashed line). Vertical arrows represent the global fractionation factors, F_N^{D-H} and F_U^{D-H} , for the folded and unfolded states, respectively. Fractionation factors represent the collective isotopic preference of all amide sites in each state relative to solvent (see Sec. 3.4.1).



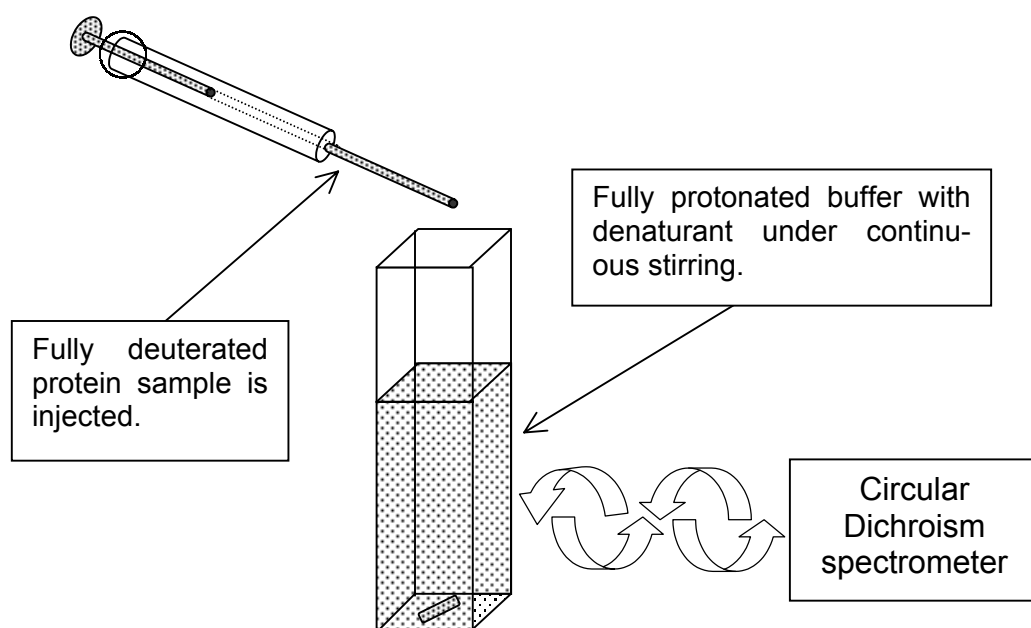
The remaining H-bonds, the carbonyl-to-water and water-to-water H-bonds, are not affected by amide substitution under constant solvent conditions and do not effect $\Delta\Delta G_{eq}^{D-H}$ as defined in Eq. 3.3

$$\Delta\Delta G^{D-H} = RT \ln \left(\frac{K_{eq}^H}{K_{eq}^D} \right) = RT \ln \left(\frac{k_f^H / k_u^H}{k_f^D / k_u^D} \right). \quad (3.3)$$

3.3.1 Experimental design

Measurement of equilibrium backbone amide isotope effects requires careful experimental design (Fig. 3.4). The effect of amide isotope substitution is determined by measuring the stability of the protein with both deuterated and protonated amides under identical solvent conditions. Fully deuterated protein is diluted 100-fold into H₂O solvent under conditions where stability (monitored by CD signal) can be determined prior to significant backbone amide exchange, pH 4.5, 10 °C (Connelly *et al.*, 1993). Here, the molecules fold within seconds and equilibrate to the new solvent condition; and the stability of the deuterated form can be determined. Side chain positions become protonated in seconds (Bai *et al.*, 1993) and do not contribute to the measured change in stability. As the hour long HX process occurs, the equilibrium of the system shifts and the CD value continuously changes (Fig. 3.5). After HX is complete, the stability of the now protonated protein is measured. The initial CD reading is used to derive the stability of the deuterated protein, K_{eq}^D , while the final reading identifies the stability of the protonated protein, K_{eq}^H , the protonated equilibrium constant. The ratio of these two

Figure 3.4 Experimental design of the equilibrium isotope effect measurements. The “ C_m experiment” (Krantz *et al.*, 2000) allows for the maximal sensitivity to the isotope effect perturbation under constant solvent conditions. Fully deuterated protein sample is injected into the 1 cm path length cuvette at a 100-fold dilution into protonated buffer with enough denaturant to unfold one-half of the protein molecules. The buffer pH and temperature are low (pH 4.5 and 10 °C, respectively) such that the rate of hydrogen exchange (HX) is slow (on the order of 500 seconds). Thus as long as the protein folding rate is faster than the HX rate, then the change of stability in the protein due to backbone amide isotopic composition can be continuously monitored by far UV circular dichroism or intrinsic fluorescence.



equilibrium constants, measured in the same bulk solvent, is used to evaluate $\Delta\Delta G_{eq}^{D-H}$ (Eq. 3.3). Additionally, measurements are conducted near the midpoint, or C_m , of the denaturation profile where the effects of a perturbation such as isotopic substitution can be accurately measured.

3.3.2 Coiled coil results

For cross-linked and dimeric GCN4, the value for $\Delta\Delta G_{eq}^{D-H}$ is ~ 0.5 kcal mol⁻¹ (Figs. 3.5a and 3.7a,b). Values of $\Delta\Delta G_{eq}^{D-H}$ are determined at several different GdmHCl concentrations and the m^o -values do not depend on whether the backbone was protonated or deuterated (Table 3.2). The C_m measurement is further validated by conducting the exchange in reverse direction, i.e. exchanging ^{NH}GCN4 into fully deuterated GdmDCl buffer (Fig. 3.5a). We find that for cross-linked GCN4, $\Delta\Delta G_{eq}^{D-H}$ (-0.37 ± 0.02 kcal mol⁻¹) and $\Delta\Delta G_{eq}^{H-D}$ (0.40 ± 0.02 kcal mol⁻¹) are equal in magnitude while opposite in direction, that is ^{NH}GCN4 loses stability upon amide exchange in D₂O whereas NDGCN4 gains stability upon exchange in H₂O.

The magnitude of the $\Delta\Delta G_{eq}^{D-H}$ value for dimeric system, -0.60 ± 0.03 kcal mol⁻¹, is greater than the cross-linked system by ~ 0.2 kcal mol⁻¹ (Table 3.2). This apparent discrepancy may be due to structural difference in the tethered region at the N-terminus as the cross-linked version's Cys-Gly-Gly tether is unable to cap the N-terminal backbone amides of Arg 1 through Gln 4. Alternatively, the melting midpoints of cross-linked and dimeric GCN4 differ by 3 M (Fig. 3.7b,c), and a slight denaturant dependence of the isotope effect may exist.

Figure 3.5 Equilibrium isotope effect time courses for α -helical proteins. The initial and final spectroscopic values for each hydrogen exchange (HX) time course determine the equilibrium stabilities, K^H (for protonated protein) and K^D (for deuterated protein). Unless otherwise noted, buffer conditions are 5 to 20 mM sodium acetate, pH 4.5 (or pD_{read} 4.1) at 10 °C. Note that the symbol conventions given for GCN4 are used throughout. **(a)** GCN4_{D7A}^{Nx} (4 μ M) in 6.0 M GdmHCl or GdmDCl is monitored by CD at 222 nm. The deuterated protein, NDGCN4_{D7A}^{Nx}, gains stability (decreasing CD signal) when exchanged into H₂O (\square) whereas ^{NH}GCN4_{D7A}^{Nx} loses stability (increasing CD signal) when exchanged into D₂O (\triangle) where exchange rates are slightly slower (Connelly *et al.*, 1993). The control experiment, ^{NH}GCN4_{D7A}^{Nx} diluted into H₂O buffer (\bullet), shows that the CD signal is constant. Native cross-linked GCN4 has -86 mdeg of ellipticity. Predicted curves ($--$) calculated according to Eq. 3.4 deviate from the data by 50% in rate, and the calculated traces may reflect a change in the predicted intrinsic rates in the presence of high concentrations of GdmHCl (Loftus *et al.*, 1986), or the assumption that the exchange of each proton has an equal contribution to $\Delta\Delta G_{eq}^{D-H}$. **(b)** GCN4_{D7A}^d (7.8 μ M) HX is monitored in 2.8 M GdmHCl where native protein has -84 mdeg of ellipticity. **(c)** Cyt *c* (5 μ M) in 2.5 M GdmHCl is monitored by Trp59 fluorescence quenching by Förster energy-transfer to the heme (Sosnick *et al.*, 1996b). When NDCyt *c* is diluted into H₂O solvent, proline isomerization and HX kinetics occur simultaneously. In order to deconvolute the proline effect, ^{NH}Cyt *c* is also diluted into the same H₂O solvent to determine the small energy change due to proline isomerization (\bullet). λ^{AA} **(h)** is affected similarly but not GCN4 **(a,b)**, since the latter does not contain proline residues. Fluorescence data are normalized to fraction unfolded as determined from signals of the fully unfolded and folded protein. **(d)** Cyt *c*^{pWT} (5 μ M) is studied by CD at 2.6 M GdmCl. Arc **(e)** and Arc^{MYL} **(f)** are studied at 10 μ M in 0.5 and 3.0 M Urea respectively. λ^{GG} **(g)** and λ^{AA} **(h)** are studied at 2.0 and 2.6M GdmCl, respectively. For IM7 **(i)**, a unique experimental design was employed, because low pH affects the protein's structure and stability adversely. A stopped-flow was used at pH 7.0 in 50 mM phosphate whereby the unfolding phase in 3.7 M Urea was deconvoluted from the C_m experiment's HX phase by a multi-exponential fit to the fluorescence data. Note the equal and opposite effect **(a,d,f,g)** when the C_m experiment is conducted in reverse (compare \square and \triangle). For equilibrium values calculated from these time courses see Table 3.2.

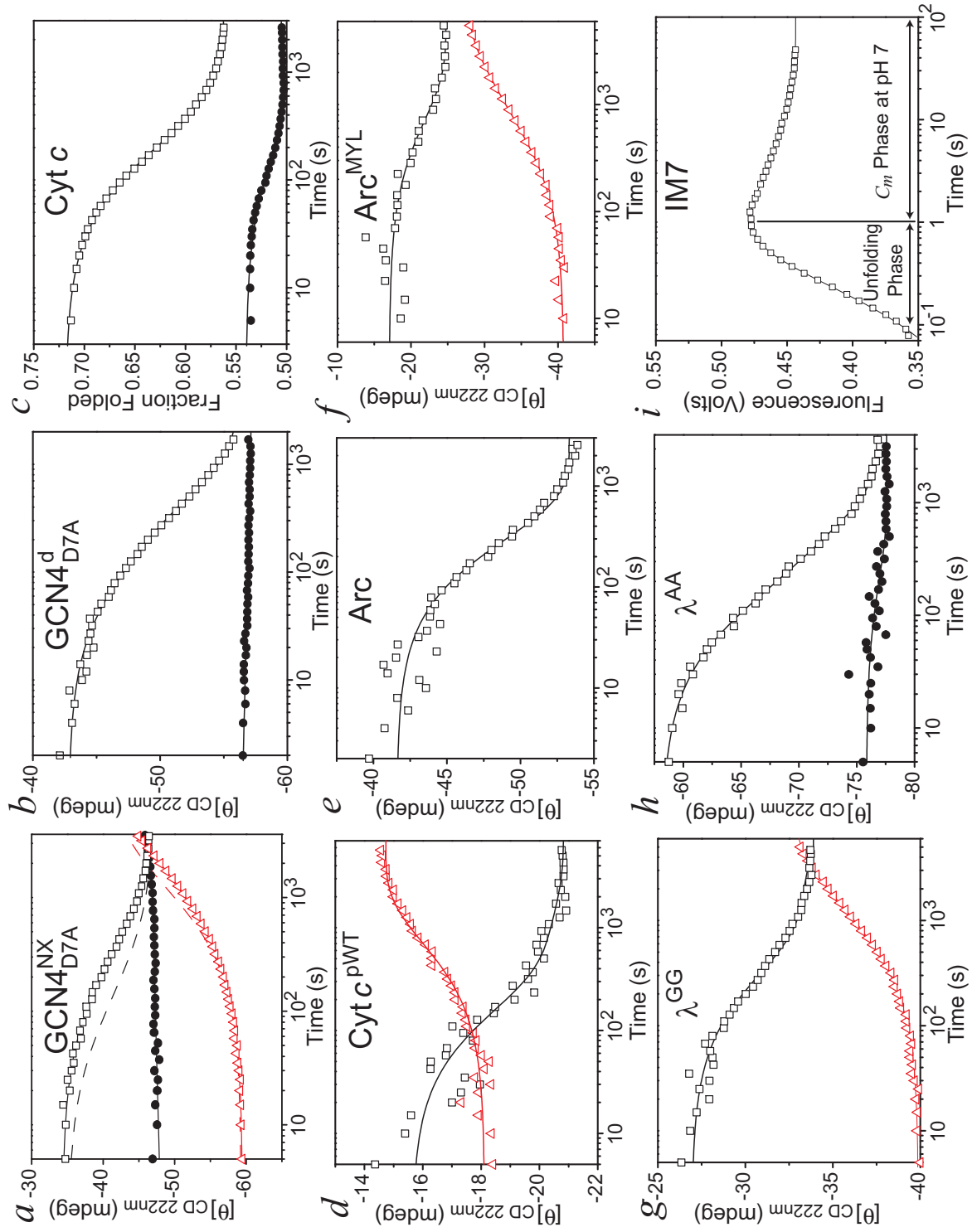


Figure 3.6 Equilibrium isotope effect time courses for β -sheet and α/β proteins. See legend of Figs. 3.4 and 3.5 for a more detailed description of the C_m experiment. Because the magnitude of the equilibrium isotope effect is smaller for largely β -sheet containing proteins, a separate pre-incubation procedure was devised to eliminate the effects of proline isomerization. Deuterated protein is pre-equilibrated in deuterated buffer to allow isomerization to reach equilibrium prior to diluting protein into protonated buffer at the same denaturant concentration. **(a)** $^{ND}CI2$ (5 μ M) is exchanged in 4 M GdmCl buffer (\square) which is fully protonated. **(b)** ^{ND}Ub (5 μ M) is exchanged into protonated 3.7 M GdmCl (\square) and a control experiment where ^{NH}Ub , pre-equilibrated to eliminate proline isomerization (see above) is diluted in protonated buffer (\bullet). Native Ub has a CD value of -16 mdeg. ProtG **(c)**, ProtL **(d)**, and SH3 **(e)** were studied similarly at (5 μ M) in 2.9 M, 2.9 M and 3.0 M GdmCl, respectively. Deuterated protein is exchanged into protonated buffer (\square) and protonated protein is exchanged into deuterated buffer (\triangle) An equal yet opposite change in stability is realized for these proteins when the HX is conducted in reverse thereby confirming these small isotope effects. Note that Ub and *src*-SH3 **(b,e)** unlike all other proteins studied have more stability when deuterated. For equilibrium values calculated from these time courses see Table 3.2.

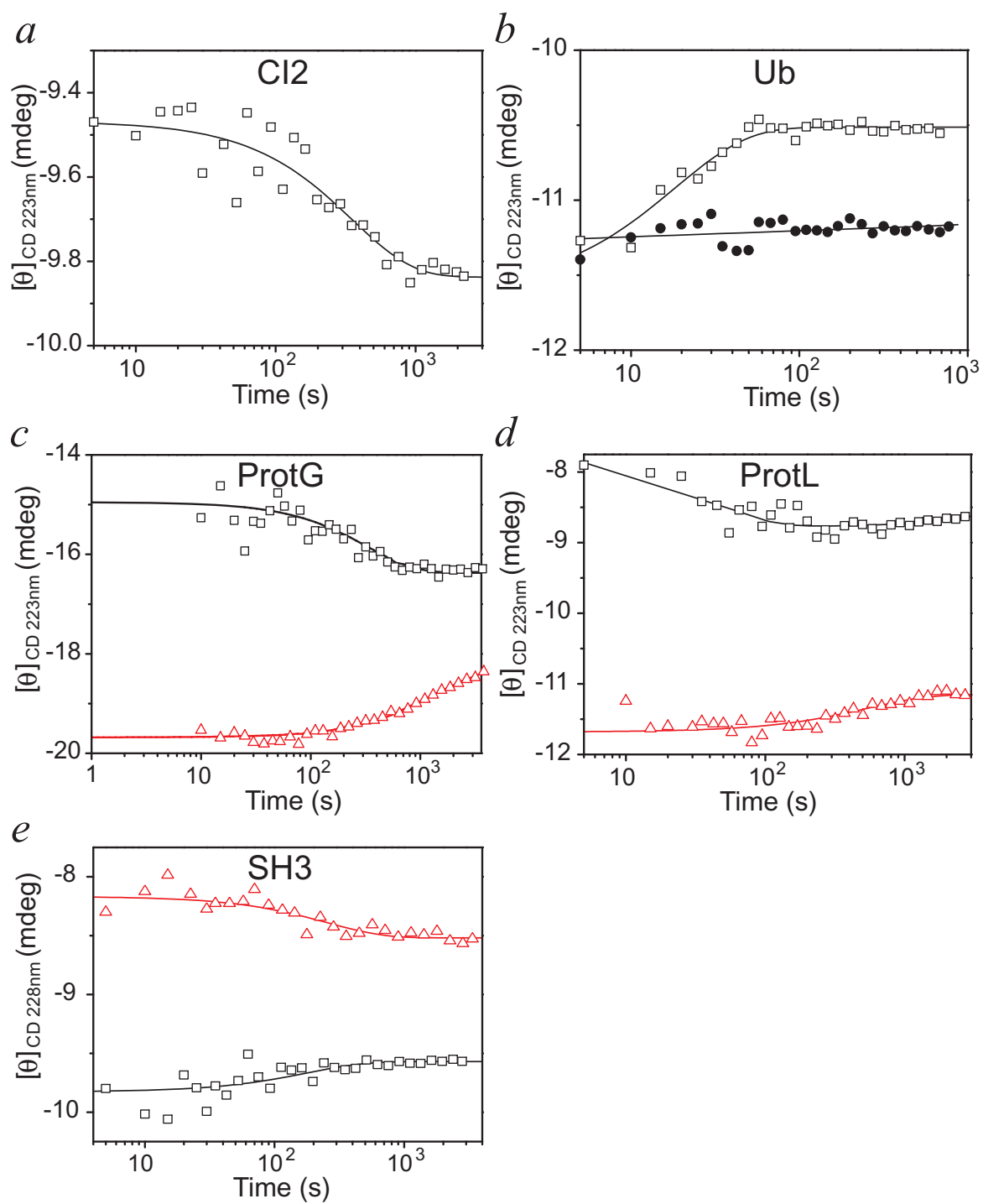


Figure 3.7 Denaturant dependence of the equilibrium isotope effects. (a) C_m experiment CD values for protonated (\blacktriangle) and deuterated protein (\bullet) obtained at multiple denaturant concentrations (see Figs. 3.5 and 3.6) can be plotted on equilibrium curves. Thus, K^H and K^D are calculated directly and may be used to determine the denaturant dependence of free energies, ΔG , in various combinations of solvent and amide composition for GCN4_{D7A}^{Nx} *(b)* and GCN4_{D7A}^d *(c)*. Open symbols represent the stability of protonated (Δ) and deuterated (\circ) protein in 99% protonated solvent, whereas solid symbols represent the stability of protonated (\blacktriangle) and deuterated (\bullet) protein in 99% deuterated solvent. Furthermore, because C_m experiments were conducted both in H₂O and in D₂O for GCN4_{D7A}^{Nx}, a bulk solvent isotope effect may be assessed by these methods (compare \blacktriangle and Δ ; or compare \bullet and \circ). For Cyt *c* *(d)*, the changes in stability due to proline isomerization, $\Delta\Delta G^{\text{Pro}}$, are obtained by measurements of ^{NH}Cyt *c* in protonated buffer and are subtracted from the value of $\Delta\Delta G^{\text{D-H/Pro}}$ for NDCyt *c* to produce the desired, proline-corrected data (\blacksquare fit with solid line). Denaturant dependence data were also estimated for λ^{AA} *(e)* and IM7 *(f)*, where the latter are plotted (\blacksquare) simply as the equilibrium isotope effect, $\Delta\Delta G_{eq}^{\text{D-H}}$, using the stopped-flow method at pH 7 outlined in Fig. 3.5*i*. The mean free energy difference between the data points for each protein studied is the $\Delta\Delta G_{eq}^{\text{D-H}}$ value presented in Table 3.2.

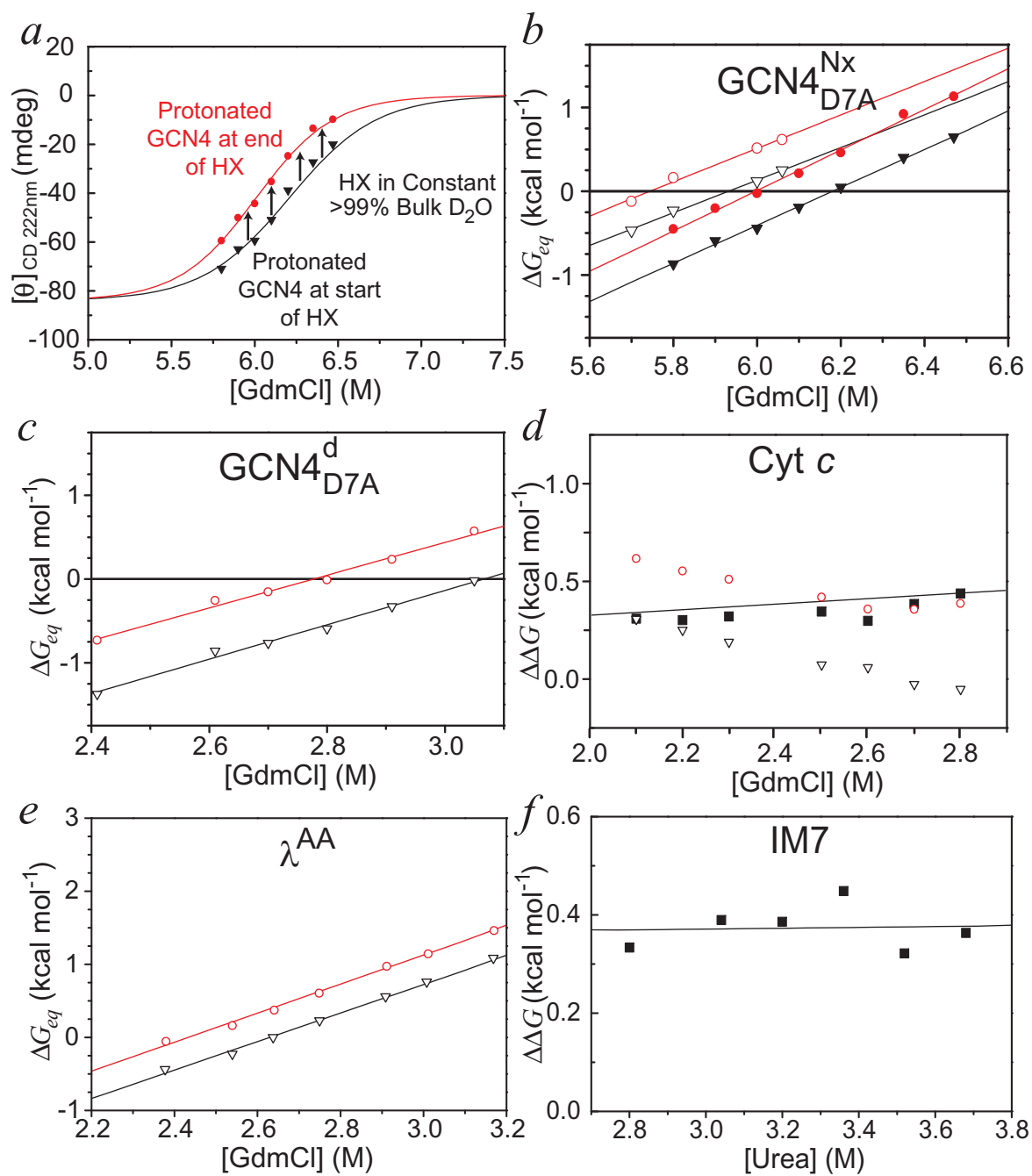


Table 3.2 Equilibrium^a and kinetic backbone amide isotope effects^b parameters

Protein	Condition (den./pH)	$-\Delta G_{Kinetics}^{H_2O}$ (kcal mol ⁻¹) ^b	$RT \ln k_f^{H_2O}$ (kcal mol ⁻¹) ^b	$m_{kinetics}^0$ (kcal mol ⁻¹ M ⁻¹) ^b	$-m_f$ (kcal mol ⁻¹ M ⁻¹) ^b	$\Delta \Delta G_{kinetics}^{D-H}$ (kcal mol ⁻¹)	$\Delta \Delta G_{eq}^{D-H}$ (kcal mol ⁻¹) ^a	ϕ_f^{D-H}	m_f/m^0
Arc	Urea / 6 ^c	1.71±0.06 (1.05±0.05)	1.89±0.03 (1.42±0.03)	1.53±0.13 ^e	1.16±0.13 ^e	-0.65±0.08	-0.47±0.03 ^f	0.72±0.08	0.76±0.02
ArcMYL	Urea / 5 ^c	4.32±0.06 (3.86 ±0.13)	2.98±0.03 (2.82±0.03)	0.95±0.02 ^e	0.38±0.02 ^e	-0.46±0.06	-0.53±0.02 ^f	0.34±0.07	0.40±0.01
CI2	GdmCl / 4.5	7.31±0.15 (7.13±0.15)	1.56±0.02 (1.40±0.02)	1.92±0.03 (1.89±0.03)	1.15±0.01 (1.13±0.01)	-0.16±0.04	-0.11±0.01	0.88±0.28	0.60±0.01
CTACP	GdmCl / 4.5	4.71±0.11 (4.37±0.10)	0.312±0.03 (0.115±0.03)	2.65±0.04 ^e	1.92±0.04 ^e	-0.33±0.04	N.D.	0.59±0.15	0.72±0.02
CTACP	Urea / 4.5	5.60±0.20 (5.30±0.23)	0.254±0.02 (0.052±0.02)	1.76±0.03 (1.82±0.05)	1.23±0.01 (1.28±0.02)	-0.40±0.03	N.D.	0.50±0.08	0.70±0.04
Cyt c	GdmCl / 4.5 ^h	6.81±0.12 (6.59±0.11)	4.07±0.04 (3.86±0.04)	2.36±0.04 (2.41±0.04)	1.17±0.03 (1.16±0.02)	-0.46±0.03	-0.34±0.05	0.41±0.03	0.50±0.01
GCN4 ^{Cx} _{biH3,7}	GdmCl / 7	5.05±0.05 (4.62±0.05)	5.85±0.04 (5.59±0.04)	1.46±0.02 ^e	1.10±0.01 ^e	-0.43±0.04	N.D.	0.61±0.03	0.75±0.02
GCN4 ^d _{D7A}	GdmCl / 4.5 ^g	12.5±0.11 (11.3±0.14)	8.89±0.10 (8.20±0.13)	1.55±0.07 (1.70±0.10)	0.82±0.06 (1.03±0.10)	-0.69±0.03	-0.60±0.03	0.59±0.02	0.53±0.05
GCN4 ^{Nx} _{D7A}	GdmCl / 4.5	13.2±0.09 (12.1±0.37)	7.21±0.09 (7.02±0.09)	2.14±0.06 (2.07±0.04)	1.08±0.02 (1.08±0.02)	-0.41±0.04	-0.37±0.02 (-0.40±0.02)	0.59±0.05	0.51±0.02
IM7	Urea / 6	2.45±0.04 (2.19±0.04)	3.17±0.02 (3.00±0.02)	0.87±0.01 ^e	0.65±0.02 ^e	-0.26±0.02	N.D.	0.62±0.04	0.75±0.02
IM7	Urea / 7	4.10±0.04 (3.60±0.04)	3.97±0.02 (3.63±0.02)	1.20±0.01 ^e	1.01±0.01 ^e	-0.50±0.03	-0.37±0.02	0.69±0.02	0.85±0.01
λAA	GdmCl / 4.5	6.72±0.17 (6.13±0.18) ^d	6.68±0.20 (6.22±0.20) ^d	2.51±0.08 ^{d,e}	1.94±0.08 ^{d,e}	-0.59±0.09	-0.39±0.01	0.78±0.05	0.76±0.04

Table 3.2 Equilibrium^a and kinetic backbone amide isotope effects^b parameters (continued)

Protein	Condition (den./pH)	$-\Delta G_{Kinetics}^{H_2O}$ (kcal mol ⁻¹) ^b	$RT \ln k_f^{H_2O}$ (kcal mol ⁻¹) ^b	$m_{kinetics}^o$ (kcal mol ⁻¹ M ⁻¹) ^b	$-m_f$ (kcal mol ⁻¹ M ⁻¹) ^b	$\Delta \Delta G_{kinetics}^{D-H}$ (kcal mol ⁻¹)	$\Delta \Delta G_{eq}^{D-H}$ (kcal mol ⁻¹) ^a	ϕ_f^{D-H}	m_f/m^o
λ AA	GdmCl / 7	5.86±0.06 (5.49±0.05)	5.41±0.05 (5.10±0.05)	2.25±0.02 ^e	1.51±0.03 ^e	-0.36±0.02	N.D.	0.84±0.04	0.67±0.01
λ AA	Urea / 7	6.38±0.13 (5.99±0.12)	5.72±0.05 (5.44±0.05)	1.12±0.02 ^e	0.84±0.01 ^e	-0.39±0.03	N.D.	0.73±0.04	0.76±0.03
λ GG	GdmCl / 7	4.16±0.04 (3.81±0.04)	4.67±0.04 (4.40±0.04)	2.16±0.03 ^e	1.53±0.03 ^e	-0.34±0.02	-0.32±0.02 ^f (0.37±0.01) ^f	0.80±0.03	0.71±0.02
λ GG	Urea / 7	5.23±0.05 (4.79±0.05)	4.83±0.03 (4.53±0.02)	1.14±0.01 ^e	0.78±0.01 ^e	-0.43±0.02	N.D.	0.69±0.02	0.68±0.01
ProtG	GdmCl / 4.5	4.84±0.03 (4.76±0.02)	3.35±0.01 (3.28±0.01)	1.84±0.01 ^e	1.57±0.01 ^e	-0.08±0.03	-0.10±0.02 (0.10±0.01)	0.79±0.3	0.85±0.01
ProtL	GdmCl / 4.5	4.89±0.02 (4.84±0.02)	1.83±0.02 (1.77±0.02)	1.90±0.01 ^e	1.34±0.01 ^e	-0.05±0.03	-0.04±0.02 (0.06±0.01)	1.1±0.8	0.71±0.01
SH3	GdmCl / 4.5	3.08±0.03 (3.16±0.03)	1.68±0.03 (1.73±0.03)	1.48±0.02 ^e	1.05±0.02 ^e	0.08±0.04	-0.09±0.02 (0.08±0.02)	0.55±0.49	0.71±0.02
Ub	GdmCl / 4.5	7.29±0.07 (7.32±0.06)	3.21±0.06 (3.19±0.04)	2.04±0.02 (2.02±0.02)	1.31±0.02 (1.30±0.02)	0.10±0.02	0.07±0.01	0.01±0.20	0.64±0.01

^a C_m experiment, which determines $\Delta \Delta G_{Eq}^{D-H}$, upon exchange of deuterated protein in constant, 99% bulk H₂O solvent. Values in

parentheses are for C_m experiments conducted in reverse (i.e. constant, 99% bulk D₂O). The sign of $\Delta \Delta G_{eq}^{D-H}$ designates the difference in free energy between that starting and ending states.

^bStopped-flow fluorescence kinetic measurements are conducted in constant ~90% H₂O solvent. Titrator fluorescence kinetics experiments are conducted in constant ~99% H₂O solvent. $\Delta G_{Kinetics}^{H_2O}$ and $RT \ln k_f^{H_2O}$ are extrapolated to 0 M denaturant, whereas $\Delta \Delta G_{kinetics}^{D-H}$ and ϕ_f^{D-H} are calculated from the middle of folding and unfolding arms of chevrons to minimize

Table 3.2 Equilibrium^a and kinetic backbone amide isotope effects^b parameters (continued)

errors of extrapolation to 0 M denaturant. Values in parentheses for kinetics parameters, $\Delta G_{Kinetics}^{H_2O}$, $RT \ln k_f^{H_2O}$, $m_{kinetics}^0$, and m_f , are for deuterated protein.

^cFree energy parameters of dimeric Arc^{MYL} and Arc are fit to data at experimental condition of 9.6 μ M and 9.3 μ M monomer concentration, respectively.

^dStopped-flow CD at 222 nm probed the kinetics of ND λ^{AA} at pH 4.5, since local HX events proximal to the chromophore potentially created fluorescence artifacts at pH 4.5. Fluorescence problems, however, were resolved when ND λ^{AA} folding was measured at pH 7.

^e m -values fit globally to deuterated and protonated kinetics data.

^f C_m experiment conducted at pH 4.5 instead of listed experimental pH.

^gThe listed values for $\Delta G_{Equil}^{H_2O}$, $\Delta G_{Kinetics}^{H_2O}$, and $RT \ln k_f^{H_2O}$ for GCN4^d_{D7A} are referenced to 1 M standard state peptide concentration. Equilibrium and kinetic experiments were conducted at protein concentrations of 7.8 μ M and 16 μ M, respectively.

^hCyt *c* experiments were conducted in 20 mM sodium azide to eliminate non-native His-heme mis-ligation.

3.3.3 The time dependence of hydrogen exchange

The time course of HX can be predicted from the sequence-dependent amide exchange rates for those protons involved in H-bonds (Bai *et al.*, 1993; Fig. 3.5a). HX essentially only occurs when the protein is unfolded. Hence, the observed exchange rate is the product of the time-dependent unfolded fraction, $f_u(t)$, obtained from the fractional CD or fluorescence signal and the random coil exchange rate, k^{rc} , for each of the n amides involved in intra-molecular H-bonds. Accordingly, the predicted CD trace can be calculated as the sum of individual exponential decays

$$\Theta_{obs}^{HX}(t) = \Theta_{final} + \frac{\Theta_{initial} - \Theta_{final}}{n} \sum_{i=1}^n \exp(-f_u(t) \times k_i^{rc} \times t). \quad (3.4)$$

The similarity between the traces calculated from Eq. 3.4 and the observed time courses confirms that the measurement is assaying for the change in stability resulting from amide HX at multiple sites throughout the protein.

3.3.4 The correlation with helical proteins

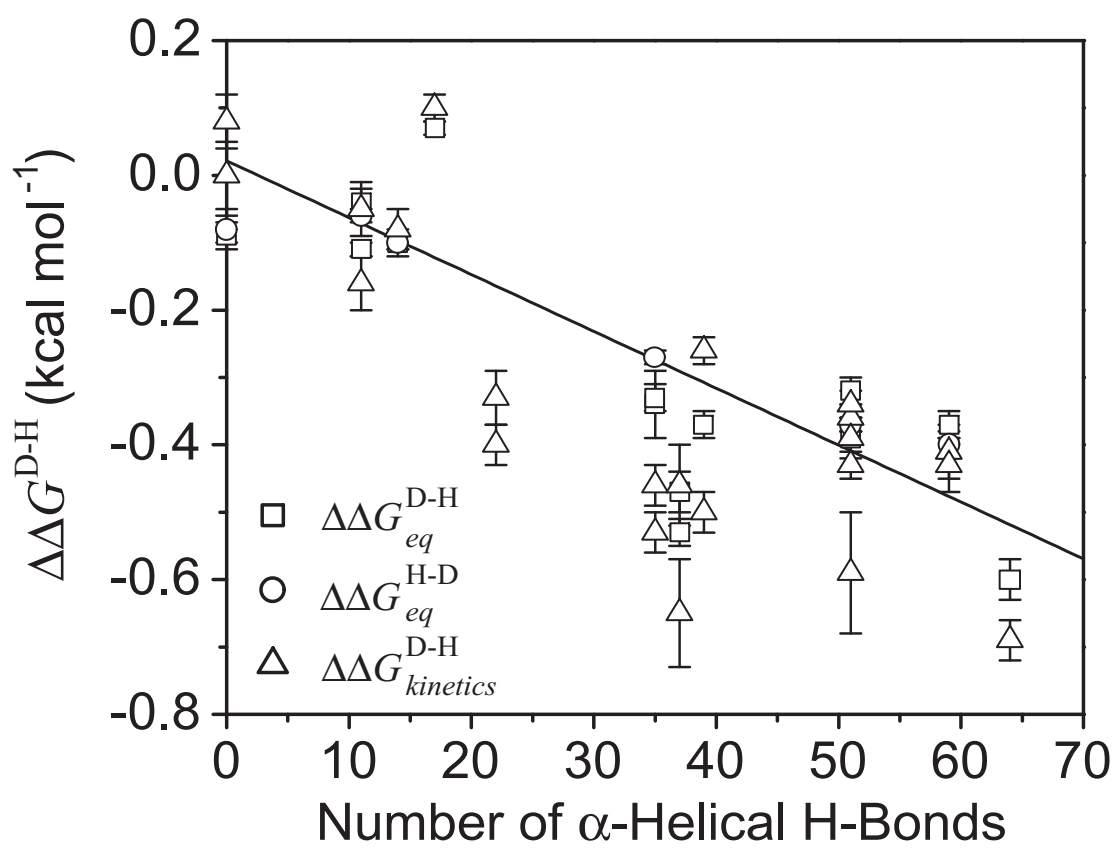
Upon establishing the existence of an equilibrium isotope effect for the coiled coil, additional C_m measurements were made for the library of the protein model systems with different secondary structure compositions (Fig. 3.2). While C_m measurements are, however, limited to those proteins in which the folding equilibrium is established faster than HX, kinetic studies (which will be described in Sec. 3.5, Figs. 3.10 and 3.11) also assess equilibrium isotope effects, $\Delta\Delta G^{D-H}$, according to a general two-state relationship (Eq. 3.3). Equality between two independently derived values for $\Delta\Delta G^{D-H}$ provides a stringent internal control for the measurement (Table 3.2).

In general, it is found that hydrogen is more stabilizing than deuterium, and the isotope effect is largest in predominantly helical proteins. These observations prompted analysis of the correlation between $\Delta\Delta G^{\text{D-H}}$ and the number of helical H-bonds in each protein (Fig 3.8). The correlation solely depends on the number of helical H-bond with a slope of $-8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ H-bond}^{-1}$. Because the y-intercept is nearly zero, and the residuals do not systematically deviate in the correlation, β -sheet H-bonds appear to have little or no isotope effect. Indeed, the isotope effect found for the all β -sheet proteins, SH3 and CD2 (Parker & Clarke, 1997), was marginal in the former and non-existent in the latter case (see also Sec. 3.4.3).

3.3.5 Solvent isotope effects

Since we have identified the effect of amide substitution on stability in cross-linked GCN4 in both H_2O and D_2O , we can isolate the energetic effect of the change in bulk solvent, $\Delta\Delta G_{eq}^{\text{D}_2\text{O-H}_2\text{O}}$. This value is $0.54 \pm 0.02 \text{ kcal mol}^{-1}$ for $^{\text{NH}}\text{GCN4}$ and $0.50 \pm 0.03 \text{ kcal mol}^{-1}$ for $^{\text{ND}}\text{GCN4}$, where D_2O is more stabilizing than H_2O solvent and is independent of the amide composition (Fig. 3.7b). These values are self-consistent and suggest that auxiliary isotope effects between protein amides and solvent molecules do not exist. That is, whether the protein possesses a fully deuterated or protonated backbone, the effect of solvent composition on stability is identical. This observation is crucial, since thermodynamic experiments utilize a bulk solvent of $\sim 99\%$ H_2O whereas the kinetic experiments use 88% H_2O . A comparison of $^{\text{NH}}\text{GCN4}$ in H_2O to $^{\text{ND}}\text{GCN4}$ in D_2O reveals a nearly zero free energy change attributable to a cancellation of the bulk

Figure 3.8 Equilibrium isotope effect correlates with helical H-bond composition. The slope, $-8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ H-bond}^{-1}$, for the least squares linear fit (**—**) is the average isotope effect per helical H-bond. The fit has an intercept of $-13 \pm 40 \text{ cal mol}^{-1}$ and R-value of 0.85.



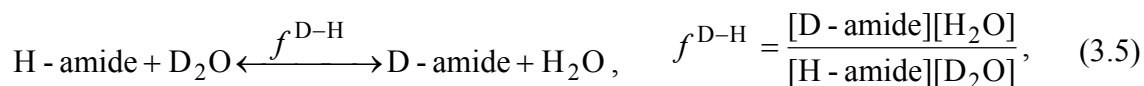
solvent effect ($0.5 \text{ kcal mol}^{-1}$) with the amide isotope effect ($-0.4 \text{ kcal mol}^{-1}$). For the all β protein, CD2, Parker and Clarke determined a value of $\sim 1.3 \text{ kcal mol}^{-1}$ for $\Delta\Delta G_{eq}^{D_2O-H_2O}$, where D_2O is more stabilizing than H_2O (Parker & Clarke, 1997).

3.4 Estimates of H-bond strength from isotope effects

From these equilibrium isotope effect results, an estimate of the free energy of the amide-carbonyl half reaction ($NH:O=\angle$) was made (Shi *et al.*, 2002). In brief, NMR measurements of chemical shifts assessed the relative preference of an internal H-bond for deuterium or protium in small molecule model compound. The preference of any H-bond for deuterium versus protium is called a fractionation factor (Bowers & Klevit, 1996; Bowers & Klevit, 2000; Edison *et al.*, 1995; Kentsis & Sosnick, 1998; Khare *et al.*, 1999; Krantz *et al.*, 2000; LiWang & Bax, 1996; Loh & Markley, 1994). Fractionation factors and the isotope effect described herein form adjacent arms in a thermodynamic cycle (Fig. 3.3).

3.4.1 Fractionation factors and isotope effects

A more detailed description of the relationship of fractionation factors and equilibrium isotope effects will be given. For each backbone amide involved in an intra-protein H-bond, the exchange process is



where $f^{\text{D-H}}$ is the equilibrium constant, or fractionation factor, for each site. “Global” fractionation factors for the native and unfolded states, $F_N^{\text{D-H}}$ and $F_U^{\text{D-H}}$ respectively, are the product of all the individual fractionation factors,

$$F^{\text{D-H}} = \prod_{i=1}^n f_i^{\text{D-H}}. \quad (3.6)$$

This estimation of global fractionation factors assumes that each individual amide site’s preference for deuterium is independent of all other sites. The global fractionation factors reflect the protein’s overall preference for deuterium in either the N and U states relative to water. These global equilibrium constants describe two arms of the thermodynamic cycle alongside the two amide isotope effect equilibrium constants (K^{D} and K^{H}) observed in the C_m experiment (Fig. 3.3; Hvidt & Nielsen, 1966). The energetic term for the equilibrium amide isotope effects observed in the C_m experiment can be written in terms of the ratio of global fractionation factors in the native and unfolded states

$$\Delta\Delta G_{eq}^{\text{D-H}} = RT \ln \left(\frac{K_{eq}^{\text{D}}}{K_{eq}^{\text{H}}} \right) = RT \ln \left(\frac{F_N^{\text{D-H}}}{F_U^{\text{D-H}}} \right). \quad (3.7)$$

Site-resolved fractionation factors in folded proteins have been observed in the range of about 0.8-1.2 with amides in α -helices tending to have values lower than those in β -sheets (Bowers & Klevit, 1996; Khare *et al.*, 1999). For the helical proteins studied here, deuteration results in a destabilization of $\sim 0.5 \text{ kcal mol}^{-1}$ or $F_N^{\text{D-H}} / F_U^{\text{D-H}} \sim 0.4$. In the mixed α/β protein, deuteration results in a mild stabilization of $\sim 0.1 \text{ kcal mol}^{-1}$ or $F_N^{\text{D-H}} / F_U^{\text{D-H}} \sim 1.2$. Recalling Eq. 3.6, the average value for the ratio of the native and

unfolded state fractionation factors at each of n amide positions,

$$\langle f_N^{\text{D-H}} \rangle / \langle f_U^{\text{D-H}} \rangle = \left(F_N^{\text{D-H}} / F_U^{\text{D-H}} \right)^{1/n}, \text{ is } \sim 0.98 \text{ for the helical molecules and } \sim 1.004 \text{ for}$$

ubiquitin, an α/β protein, consistent with the trend observed for these two classes of secondary structures (Bowers & Klevit, 1996; Khare *et al.*, 1999). Site-resolved NMR methods further confirm that the fractionation factors for amides in Ub (LiWang & Bax, 1996) are greater than one and prefer deuterium over hydrogen. These results and previous results (Kentsis & Sosnick, 1998) indicate that the ratio of fractionation factors in the N and U states are near unity. Therefore, it may be inferred that amide-backbone H-bonds in the native state are equal in strength to amide-water H-bonds in the unfolded state.

3.4.2 A scale factor for H-bond strength

In order to make an order of magnitude estimate of the contribution of H-bond formation to the stability of a protein, however, I must return to the NMR study by Shi *et al.* Here, N-acetylanthranilic acid models carbonyl-amide hydrogen bonds formed during protein folding, and 4-acetamidobenzoic acid models the unfolded state in which the amide hydrogen bonds to water. NMR is used to measure shifts in $\text{p}K_a$ of the ionizable carboxyl group when the amides of the compounds are either protonated or deuterated. From the $\text{p}K_a$ shift, a quantitative scale factor, SF, is obtained:

$$SF = \frac{\partial(\Delta G_{eq}^{\text{H-Bond}})}{\partial(RT \ln f^{\text{D-H}})}, \quad (3.8)$$

where ΔG^{HB} is the change in free energy of a hydrogen bond upon isotope substitution, and f^{D-H} is the fractionation factor of the model compound. The scale factor reported is 74 ± 27 .

Referring to Eqs. 3.5-3.8, the free energy change of the formation of a backbone H-bond during protein folding, $\Delta\Delta G_{eq}^{\text{H-Bond}}$, is derived by the simple relationship

$$\Delta\Delta G_{eq}^{\text{H-Bond}} = \Delta\Delta G_{eq}^{\text{D-H}} \times SF. \quad (3.9)$$

Because the global isotope effect for each protein studied varies widely (Table 3.2, Fig. 3.8), the contribution of H-bond formation also varies widely from -50 to $+6 \text{ kcal mol}^{-1}$.

3.4.3 α -helix versus β -sheet

The origin of these differences in H-bond stabilization may be due to the secondary structure content of the individual proteins. As indicated by the correlation in Fig. 3.8, the isotope effect, $\Delta\Delta G^{\text{D-H}}$, varies directly with α -helical content, i.e. $-8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ H-bond}^{-1}$. Thus, applying the scale factor to the average helical isotope effect, the average helical H-bond contributes $0.64 \text{ kcal mol}^{-1}$ of stabilization. Whereas the net contribution of the β -sheet H-bond in any given protein is negligible, since no systematic trend in the isotope effect per number of β -sheet H-bond sites is discernable and the intercept to Fig. 3.8 is nearly zero (see Sec. 3.3.4).

The pronounced difference between an H-bond found in α -helix versus β -sheet may be explained by several factors that include (i) the pK_a differences of the donor and acceptor pair and (ii) the intrinsic geometric differences of two secondary classes. The

intrinsic pK_a for each amide and carbonyl group can differ significantly, even for the same residue found in a different sequence (Bai *et al.*, 1993). The geometry of H-bonds are also variable, especially the angles between the N-H and O=C groups in helices, parallel and anti-parallel β -sheets. Hence, it may not be surprising to find that H-bonds contribute differently in different proteins with different secondary structures and environments.

Helical H-bonds often have a nearly homogeneous, nearly optimal bonding geometry and occupy a much more narrow portion of ϕ - ψ space on the Ramachandran map than β -sheet H-bonds. Moreover, there is an inherent degeneracy in β -sheet structures in terms of strand topology (parallel *versus* anti-parallel). While the degeneracy may seem to obfuscate this discussion (since parallel and anti-parallel H-bonds were not experimentally dissected), the isotope effect may only be present in certain regions of the Ramachandran map. Thus, the more diverse β -sheet class may, in effect, cancel favorable geometries with unfavorable geometries. Thus, it can be argued that the very wide sampling of many geometries and conformation for β -sheet H-bonds may explain in part why they lack an apparent isotope effect and consequently contribute little to the stability of the protein.

Now taking the argument in reverse, the well documented contortion and twisting of the β -stands (Branden & Tooze, 1991) may indicate that H-bond geometry is weakly constrained, because there is neither a significant energetic benefit nor penalty placed upon β -sheet H-bond geometry in the first place. Thus the determinant of β -sheet H-bond

geometry may be implicit to close hydrophobic packing of side chains rather than explicit to the electrostatic potential of the H-bond itself.

3.4.4 Limitations and implications of the scale factor result

Finally, all of these arguments admit that a significant penalty is paid when an H-bond is broken in the context of a highly desolvated, hydrophobic protein core. The discussion of “H-bond strength” is always the relative strength all H-bonds in the protein folding reaction, comparing the water-amide and water-carbonyl H-bonds in the unfolded state to the water-water and amide-carbonyl H-bonds formed in the native state. A critical limitation of the present estimation of the scale factor is that the water-carbonyl and water-water components of the reaction are not included. While Shi *et al.* argues that acetone-water experiments provide insight on the water-water/water-carbonyl component of the thermodynamic cycle, it cannot be avoided that acetone is not chemically equivalent to a true protein carbonyl.

Thus, an exact comparison is difficult to make between the present study and previous estimates of the net strength of an H-bond. As aforementioned, the stability derived from the isotope effect measurement reflects the energetic contribution exclusively from the amide-carbonyl and amide-water H-bonds, with none of the additional contributions that are normally included in other measurements such as polar desolvation. For example, Pace and coworkers estimated $\Delta G \sim -1$ to -2 kcal mol^{-1} H-bond⁻¹ from mutational studies, which includes both the loss of water-carbonyl and the gain in water-water H-bonds, in addition to the transfer of the H-bond from solution

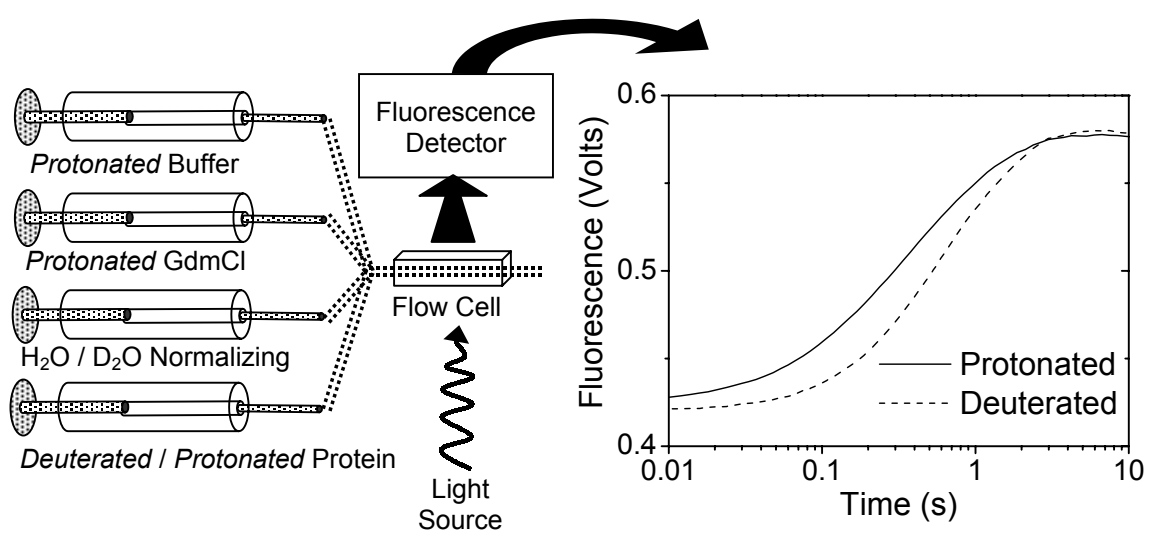
to the interior of the protein (Myers & Pace, 1996; Pace *et al.*, 1996). Similarly Ross and Rekharsky determined the equilibrium constants for a series of small organic molecules binding to cyclodextrins, concluding that at room temperature one H-bond contributes $-0.6 \text{ kcal mol}^{-1}$ to stability of the complex (Ross & Rekharsky, 1996). A number of theoretical studies arrive at the opposite conclusion (Dill, 1990; Lazaridis *et al.*, 1995; Yang & Honig, 1995), arguing that the significant penalty associated with desolvation of the polar groups offsets any gain in hydrogen bond stability.

3.5 Kinetic isotope effects

The isotope effect on folding and unfolding rates and their denaturant dependencies were obtained for the library of two-state proteins (Jackson, 1998; Fig. 3.1; Table 3.1). Implementation of a four-syringe stopped-flow protocol (Fig. 3.9) and use of the same buffers for deuterated and protonated backbone protein samples rigorously controlled isotope composition and denaturant concentration (Krantz *et al.*, 2000). For proteins with small isotope effects, $< 0.2 \text{ kcal mol}^{-1}$, we developed a simpler and more accurate method of determining rates, which takes advantage of these α/β and β proteins' fortuitously slow relaxation rates: a computer-driven titrator rapidly dilutes 100-fold either a deuterated or protonated protein into a 1 cm^2 cuvette under vigorous stirring (dead-time $\sim 0.5 \text{ s}$). The experiment is highly controlled, because the deuterated and protonated versions are alternatively injected into the same solution, thereby eliminating any possible artifacts of denaturant concentration error, pH/pD or temperature variation.

Kinetic isotope effects were obtained using the 'chevron' analysis of the denaturant dependence of the folding and unfolding activation free energies.

Figure 3.9 Four syringe protocol for a Biologic SFM4 stopped flow. This protocol maintains constant D₂O levels during the kinetic isotope effect measurements. (*left*) The solvent normalizing syringe is used to deliver either deuterated or protonated buffer solution for ^{NH}protein or NDprotein folding measurements, respectively. This protocol ensures that the D₂O and GdmHCl/GdmDCI concentration as well as buffer pH/pD are the same in the ^{NH}protein and NDprotein measurements. The three buffer solution syringes are mixed prior to the final mixing with the protein containing solution syringe. Example kinetics traces for GCN4_{D7A}^d folding kinetics (*right*). Notice that ^{NH}GCN4_{D7A}^d (—) folds faster than NDGCN4_{D7A}^d (---) by ~1.5 fold.



The denaturant dependence of these quantities, ΔG_f^\ddagger and ΔG_u^\ddagger , respectively, are linear, where the slope of the folding (m_f) and unfolding (m_u) arms relate to the amount of denaturant accessible surface area buried pre- and post-folding TS, respectively (App. 3). Importantly, these slopes are the same for the protonated and deuterated versions of each protein. This feature enables the folding and unfolding kinetic isotope effects, $\Delta\Delta G_f^{\ddagger, \text{D-H}}$ and $\Delta\Delta G_u^{\ddagger, \text{D-H}}$, to be determined from the average offset of the respective chevron arms that occurs upon isotopic substitution. We find that the equilibrium isotope effect, determined from the kinetic isotope effect measurements according to $\Delta\Delta G_{eq}^{\text{D-H}} = \Delta\Delta G_f^{\ddagger, \text{D-H}} - \Delta\Delta G_u^{\ddagger, \text{D-H}}$, corroborate the value obtained from the equilibrium C_m measurements (Table 3.2).

3.6 Analysis of kinetic isotope effects

Underlying the analysis of kinetic isotope effects is the established methodology of mutational ϕ -analysis (Fersht *et al.*, 1992; Matthews, 1987; App. 3). The effect of an amino acid substitution is quantified by the parameter ϕ_f , given by the change in folding activation free energy, $\Delta\Delta G_{fold}^\ddagger$, divided by the change in global stability, $\Delta\Delta G_{eq}$. A ϕ_f -value is the degree to which the total energetic effect of the substitution is realized in the TS. Generally, a value of zero indicates that the mutated residue is unstructured in the TS whereas a value of one is consistent with this residue sensing a native-like environment in the TS.

Here, instead of an amino acid mutation, we globally substitute all backbone amide protons with deuterons, perturbing the stability of all helical H-bonds. The isotopic $\phi_f^{\text{D-H}}$ -value is the ratio of the rate and equilibrium constants of the deuterated and protonated proteins

$$\phi_f^{\text{D-H}} = \frac{\Delta\Delta G_f^{\text{D-H}}}{\Delta\Delta G_{eq}^{\text{D-H}}} = \frac{\ln\left(k_f^{\text{D}}/k_f^{\text{H}}\right)}{\ln\left(K_{eq}^{\text{D}}/K_{eq}^{\text{H}}\right)}. \quad (3.10)$$

The interpretation of $\phi_f^{\text{D-H}}$ is more complex than for a single site ϕ_f -value, because all backbone amide sites simultaneously are altered. The linear relationship between $\Delta\Delta G_{eq}^{\text{D-H}}$ and helical H-bonds indicates that $\phi_f^{\text{D-H}}$ will only provide information on α -helical H-bonds. Thus, we equate $\phi_f^{\text{D-H}}$ with the global fraction of H-bonds formed in the TS with two assumptions. Firstly, all helical H-bond sites contribute equally to $\Delta\Delta G_{eq}^{\text{D-H}}$. Secondly, H-bonds in the TS ensemble must either be formed or broken, but they do not exist in a partially formed conformation. The all-or-none scenario is confirmed by our recent studies using engineered metal binding sites (Krantz & Sosnick, 2001) and mutational studies of the GCN4 coiled coil (Moran *et al.*, 1999). In the TS of the coiled coil, about half of the molecule is structured while the remainder is largely unstructured, with $\phi_f^{\text{D-H}} \sim 0.6$. Therefore, equating $\phi_f^{\text{D-H}}$ to the fraction of native-like H-bonds is more valid than the improbable alternative: i.e., all H-bonds simultaneously are in an energetically strained or distorted geometry in the TS with each possessing a 60% isotope effect.

Figure 3.10 Folding kinetic isotope effect measurements for α -helical proteins. Kinetic activation energy plots for **(a)** GCN4_{D7A}^{Nx} at pH 4.5, **(b)** GCN4_{D7A}^d at pH 4.5, **(c)** GCN4_{biH3,7}^{Cx} at pH 7, **(d)** Cyt *c* at pH 4.5, **(e)** Cyt *c*^{pWT} at pH 5, **(f)** Arc and Arc^{MYL} at pH 6 and 5, respectively, **(g)** λ^{AA} and λ^{GG} at pH 7, and **(h)** IM7 at pH 6 and 7. All protonated (\square or \blacksquare) and deuterated (\triangle or \blacktriangle) protein kinetic measurements are conducted at 10 °C. Standard stopped-flow methods (~1 ms dead time) are employed.

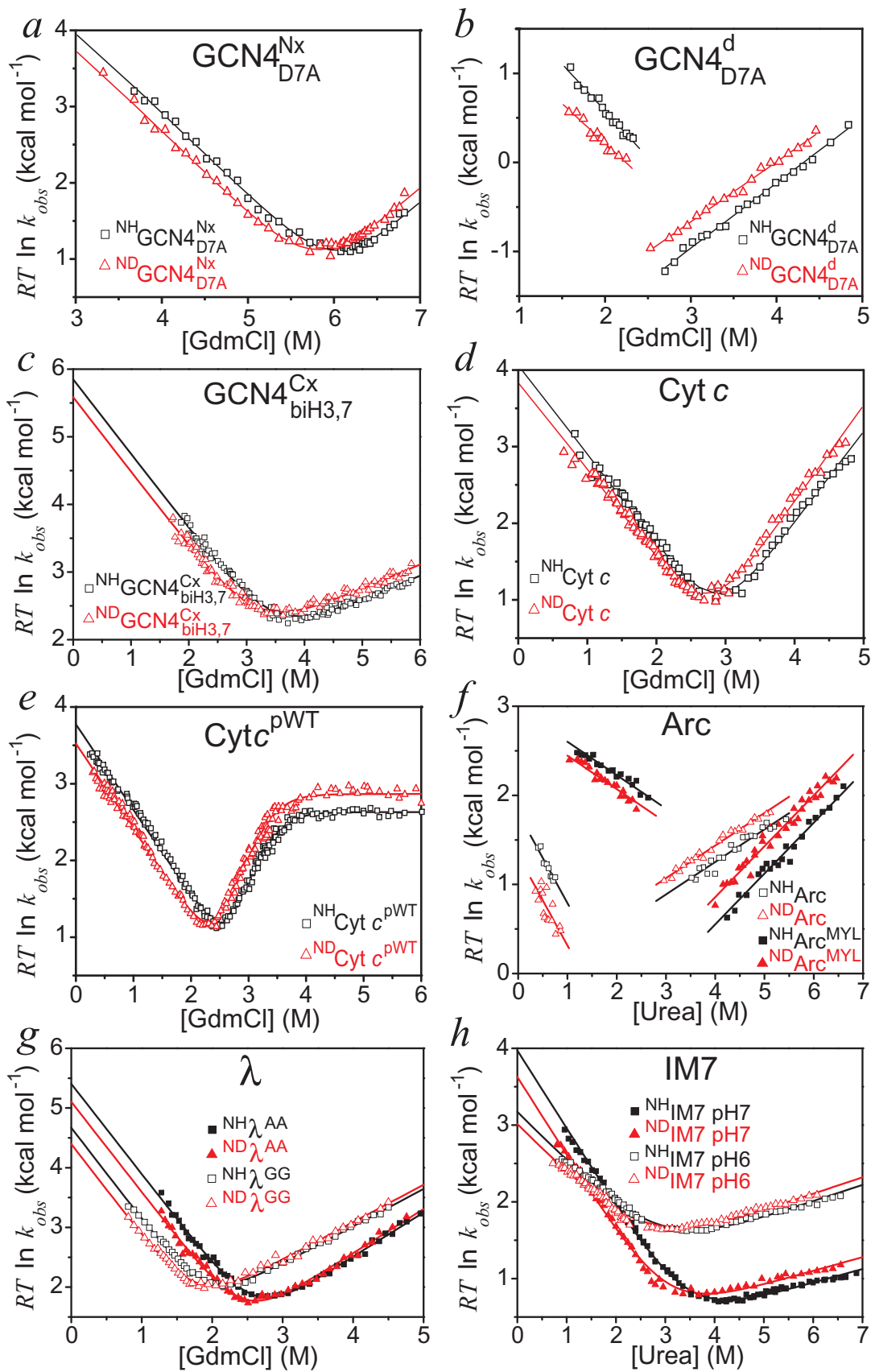
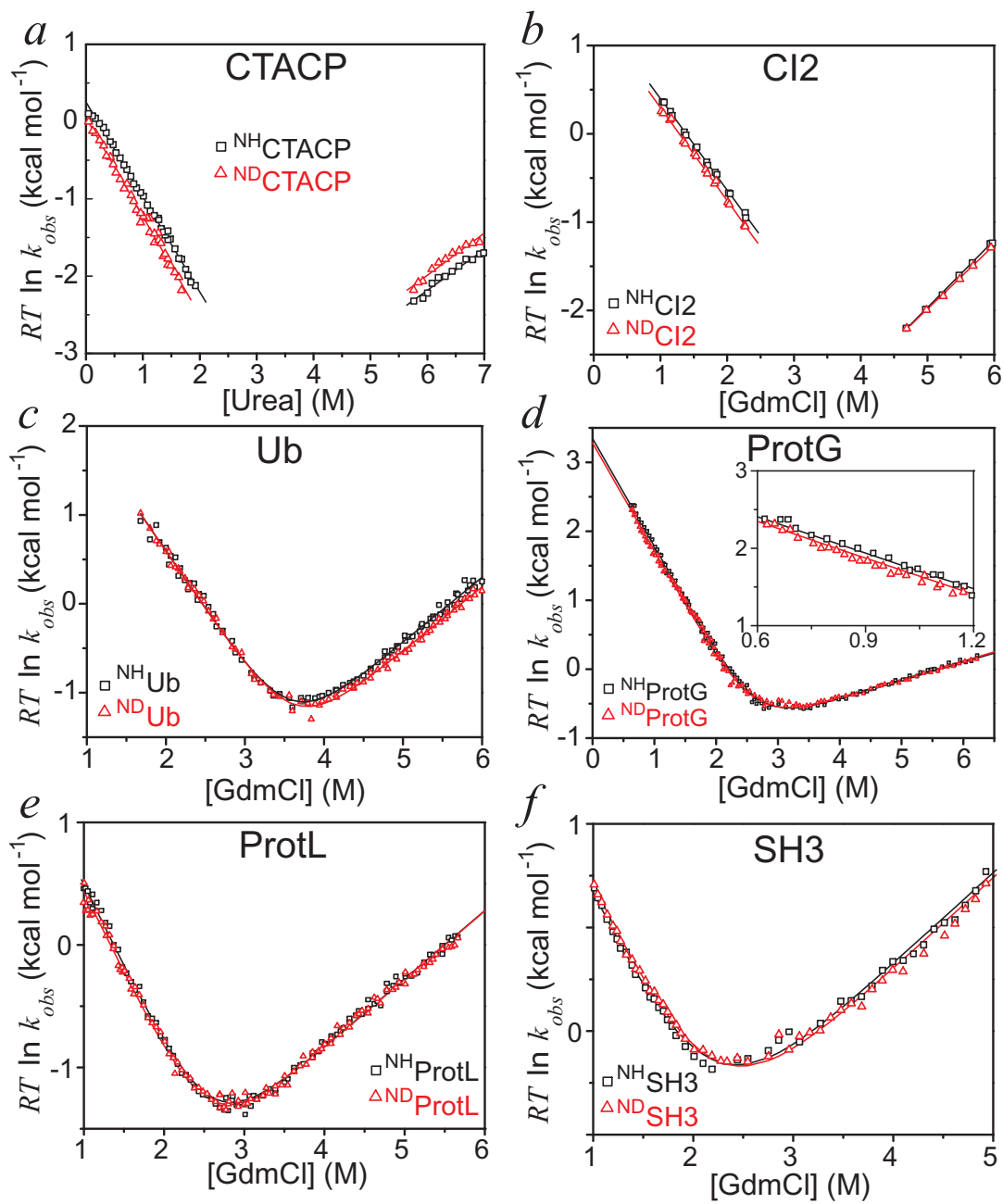


Figure 3.11 Folding kinetic isotope effect measurements for mixed α/β and β -sheet proteins. Kinetic activation energy plots for **(a)** CTACP at pH 4.5, **(b)** CI2 at pH 4.5, **(c)** Ub F45W at pH 4.5, **(d)** ProtG at pH 4.5, **(e)** ProtL at pH 4.5, **(f)** *src*-SH3 domain at pH 4.5. All protonated (\square) and deuterated (\triangle) protein kinetic measurements are conducted at 10 °C. Standard stopped-flow methods (~ 1 ms dead time) are generally employed, although these slower folding proteins, which have smaller isotope effects, are also measured by rapid dilution with an automated titrator (dead time ~ 0.5 s).



Finally, since β -sheet H-bonds are energetically insensitive to the isotopic substitution (Fig. 3.8), mixed α/β protein's folding kinetics may be analyzed with this method, but only in terms of whether α -helical H-bonds are critical to the TS ensemble.

3.7 Assessing specific protein folding models with $\phi_f^{\text{D-H}}$

3.7.1 GCN4 coiled coil

For the uniformly helical GCN4 molecules (Fig. 3.1a), $\Delta\Delta G_{eq}^{\text{D-H}}$ probably is the result of nearly equal contributions from each amide involved in the single type of H-bond. In this situation, the $\phi_f^{\text{D-H}}$ -values of 0.59 ± 0.02 and 0.59 ± 0.05 for dimeric GCN4 and N-terminally linked GCN4 equate to the percentage of H-bonds formed in the TS (Fig. 3.10a,b). GCN4_{biH3,7}^{Cx} showed similar results (Fig. 3.10c). These $\phi_f^{\text{D-H}}$ values agree with mutagenesis data of both versions of the coiled coil, which indicated 30-50% of the protein is helical in the TS (Moran *et al.*, 1999). Despite substantial pathway heterogeneity for dimeric coiled coil and a homogenous, polarized TS in cross-linked coiled coil, there remains a ~50% native strength H-bond requirement on all folding routes.

3.7.2 Cytochrome c

Cyt c (Bushnell *et al.*, 1990; Fig. 3.1b) contains 51 intra-protein H-bonds of which 35 (68%) are located in four helices and 16 in turn/coil regions (31%). If all H-bonds equally contribute to $\Delta\Delta G_{eq}^{\text{D-H}}$, the observed $\phi_f^{\text{D-H}}$ -value 0.41, equates to 21 helical H-bonds in the TS, i.e. 41% of 51 (Fig. 3.10d,e). This number can be compared

with kinetic data obtained from HX pulse labeling methods (Roder *et al.*, 1988; Sosnick *et al.*, 1996b). Cyt *c* folds in a three-state manner with a kinetic intermediate at pH 6.3 where non-native His-heme ligations are present in the denatured state. The intermediate contains the two major, N- and C-terminal helices. By comparing folding behavior under two and three-state conditions, we concluded that a similar two-helix intermediate which lacks the non-native ligation is present *after* the TS and does not accumulate under two-state conditions (Sosnick *et al.*, 1996b). In conjunction with the pulsed labeling results, our present data which indicate that 21 H-bonds are present in the TS, argues that these two helices are formed at the rate-limiting step under two-state folding conditions (Colon *et al.*, 1996; Ptitsyn, 1998; Travaglini-Allocatelli *et al.*, 1999).

3.7.3 Arc repressor

In Arc, a small dimeric transcription factor, ~75% of its backbone H-bonds are helical while 8-15% are involved in the two β -strands (4-8 H-bonds, depending upon NMR structure) which form along the dimerization interface (Fig. 3.1c). Waldburger *et al.* dissected the importance of a partially buried salt-bridge and H-bond formed by a triad of side-chains, Arg 31, Glu 36, and Arg 40 (Waldburger *et al.*, 1996). When replaced with hydrophobic side chains (R31M, E36Y, and R40L), this Arc^{MYL} variant folded 10 to 1000-fold faster, depending upon denaturant concentration. Furthermore, the surface burial in the TS shifted from $m_f/m^o \sim 0.75$ in wild type to 0.4 in Arc^{MYL}. Given our goal of determining whether helical structure formation and surface burial are concerted processes, the change in m_f/m^o makes Arc repressor an ideal comparative system to study. The $\phi_f^{\text{D-H}}$ -values for Arc and Arc^{MYL} were 0.72 ± 0.08 and 0.34 ± 0.07 ,

respectively (Fig. 3.10f). These values provide a remarkable correlation with the measured surface burial requirement of the TS (Table 3.2). Therefore, the proposal that Arc^{MYL} has an earlier TS than wild type in terms of surface burial now has the added implication that the earlier TS forms fewer H-bonds than wild-type.

Work by Srivastava and Sauer further characterized the extent to which helical structure forms in the TS of wild type Arc using Ala/Gly substitutions (Srivastava & Sauer, 2000). They found a high average ϕ_f -value (0.82 ± 0.56) for substitutions in helix B (16 H-bonds in dimer), which forms the dimerization interface, whereas the average ϕ_f -value (0.26 ± 0.22) in helix A (26 H-bonds in dimer) was significantly lower. The present isotope effect data for wild type indicates that 72% of the helical H-bonds are formed in the TS. In combination with the Ala/Gly results, we surmise that most of helix B but about half of helix A are formed in the TS. By extension, Arc^{MYL} probably has only a portion of helix B formed in the TS.

Interestingly, Arc^{MYL} buries less H-bonds in the TS than wild type but folds markedly faster (Waldburger *et al.*, 1996), indicating that more extensive helical structure formation in the TS does not necessarily result in faster folding. Furthermore, the near diffusion-limited folding rate of Arc^{MYL}, $2 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ at 25 °C, indicates that the major fraction of chain encounters are productive. Assuming that helix B is formed in the TS, as indicated by the mutational and isotope effect data, folding is two to three orders of magnitude too fast to be contingent on transient pre-formed helix B. In such a diffusion-collision scenario, k_f is only $4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$; that is, the product of the calculated 0.05%

probability of both monomers being helical (Munoz & Serrano, 1997) and the diffusion-limited encounter rate, $\sim 10^9 \text{ M}^{-1} \text{ s}^{-1}$, assuming even 100% encounter success frequency.

3.7.4 *Lambda repressor*

Burton *et al.* previously studied helix formation in the 80 residue, N-terminal domain of λ -repressor (residues 6-85, Fig. 3.1*d*) using multiple Ala/Gly substitutions (Burton *et al.*, 1997; Burton *et al.*, 1996). We choose to study two forms of this five helix bundle, wild type (λ^{GG}) and the G46A/G48A mutant form (λ^{AA}), under several solvent conditions. $\phi_f^{\text{D-H}}$ -values for λ^{GG} and λ^{AA} in urea are 0.69 ± 0.02 and 0.73 ± 0.04 , respectively, while the values are about 10% larger in GdmCl (Fig. 3.10*g*, Table 3.2). The G46A/G48A double mutation slightly reduces the extent to which the 51 α -helical H-bonds form in the TS. The amount of surface area buried in the TS is mildly reduced as well, consistent with the direction, but smaller in magnitude than observed by the NMR linewidth measurements at 37 °C (Burton *et al.*, 1997; Burton *et al.*, 1996). For both forms of λ -repressor, $\phi_f^{\text{D-H}}$ and m_f/m° are well correlated, with approximately 70-80% of the α -helical H-bonds formed. Consistently, high mutational (Ala/Gly) ϕ -values were observed in helices 1, 4 and 5 which account for 34 H-bonds, or 67% of the helical content.

3.7.5 *IM7*

The colicin immunity protein, IM7, is an 87 residue four helix bundle that has 39 backbone amides involved in α -helical H-bonds (Fig. 3.1*e*). At pH 6 and 7, the folding behavior is consistent with a two-state model, e.g. chevron arms are linear with

denaturant concentration. Also, upon reduction of protein concentration below μM , we do not observe a burst-phase fluorescence signal which is the signature of a sub-millisecond species (Capaldi *et al.*, 2001; Ferguson *et al.*, 1999). Additionally, all the H-bond formation can be accounted for in the observed reaction, as $\Delta\Delta G_{eq}^{\text{D-H}}$ is recapitulated from the kinetic measurements (Table 3.2). Finally, the folding kinetic isotope effect remains constant, which is inconsistent with the accumulation of a highly helical intermediate at lower denaturant concentrations.

Kinetic isotope effect analysis conducted at pH 7 found a $\phi_f^{\text{D-H}}$ -value of 0.69 ± 0.02 while m_f/m° is 0.85 ± 0.01 (Table 3.2). At the lower pH 6 condition, the protein buries about 1/3 less surface area (m -value) in both the TS and native state (Fig. 3.10h). Equilibrium denaturant melting experiments reproduce this effect and there is an $\sim 10\%$ decrease in the $\text{CD}_{222\text{nm}}$ signal (data not shown). Moreover, the overall equilibrium isotope effect, $\Delta\Delta G_{eq}^{\text{D-H}}$, decreased from $0.50 \text{ kcal mol}^{-1}$ to $0.26 \text{ kcal mol}^{-1}$ when the pH was lowered. Hence, the pH 6 condition alters the average surface area burial and helical H-bond structure. At pH 6, $\phi_f^{\text{D-H}} = 0.62 \pm 0.04$, and $m_f/m^\circ = 0.75 \pm 0.02$. For IM7, we conclude that approximately 70% of the helical H-bonds form at the TS.

3.7.6 CTACP

CTACP is an 11 kDa mixed α/β protein composed of five β -strands and two α -helices (Fig. 3.1g). The distribution of CTACP's 59 backbone amides is 52% in the β -sheet network, and 37% in the helices. The $\phi_f^{\text{D-H}}$ -values of 0.50 ± 0.08 and 0.59 ± 0.15

were determined in urea and GdmCl denaturants, respectively (Table 3.2, Fig. 3.11a). Because the β -sheet H-bonds are insensitive to isotopic substitution, a $\phi_f^{\text{D-H}}$ -value of 50% indicates half of the α -helical bonds are required in the TS of CTACP's folding pathway. Taddei *et al.* analyzed the folding kinetics of CTACP using site-directed mutagenesis in the two helices (Taddei *et al.*, 2000). They concluded that only helix 2 was required in the rate limiting step of the folding pathway as its ϕ_f -values were near unity, while they were near zero in the equally sized helix 1. Our data further support that only helix 2 is formed in the TS, because the global $\phi_f^{\text{D-H}}$ -value is ~ 0.5 . The kinetically important Pro 54 ($\phi_f=0.98$) in the initiating turn for helix 2 may guide formation of the β -strands and provide other crucial hydrophobic context for the helix in the TS.

3.7.7 CI2

Itzhaki *et al.* extensively studied the folding of a small α/β protein, CI2, (Fig. 3.1i) using mutational ϕ -analysis (Itzhaki *et al.*, 1995b). They determined that the ϕ_f -values were the most significant in the single helix (11 H-bonds) with an average value of 0.44 ± 0.27 . However, because very few of the helical residues had significantly high ϕ_f -values (>0.7), they concluded that the helix was either partly formed or possessed an extended geometry in the TS. Although the equilibrium isotope effect is small, $110\text{-}160 \text{ cal mol}^{-1}$, it is consistent with the 11 native α -helical H-bonds. Kinetic analysis yielded a $\phi_f^{\text{D-H}}$ -value near unity, indicating that most of the helical H-bonds are formed in the TS (Fig. 3.11b, Table 3.2).

3.7.8 Ubiquitin

Ub (Vijay-Kumar *et al.*, 1987; Fig. 3.1f) contains 45 intra-protein H-bonds involving backbone amides of which 17 are in a short 3_{10} -helix and a larger α -helix (37%), 19 are in four strands of β -sheet (42%), and 9 are in various turn/coil structures (19%). The energetic effect of amide substitution for Ub is smaller than in the α -helical proteins, but native Ub prefers D over H as demonstrated by NMR (LiWang & Bax, 1996). Perhaps these results reflect underlying properties of β -sheet containing proteins and represent the limitation of the method. The zero $\phi_f^{\text{D-H}}$ -value, however, is suggestive that H-bonds are largely absent in the TS (Table 3.2; Fig. 3.11c). Certainly, corroborating data are needed to support this explanation for Ub's kinetic isotope effect, although this interpretation is consistent with recent studies on other experimental (Grantcharova *et al.*, 1998; Gruebele & Wolynes, 1998; Kim *et al.*, 1998; Martinez *et al.*, 1998; Munoz *et al.*, 1998; Schonbrunner *et al.*, 1997) and theoretical (Guo & Thirumalai, 1997; Honeycutt & Thirumalai, 1992; Matheson & Scheraga, 1978; Nymeyer *et al.*, 1998) β -sheet systems.

3.7.9 Protein G

Protein G has 35 backbone amides involved in H-bonds of which 14 are involved in a single α -helix while the remaining are involved in a four stranded β -sheet network (Fig. 3.1h). McCallister *et al.* determined that the average non-negative, mutational ϕ_f -value in the α -helix was 0.25 ± 0.13 (McCallister *et al.*, 2000). The $\phi_f^{\text{D-H}}$ -value observed for the isotope effect is 0.79 ± 0.30 , which suggests a larger proportion of the α -helix is

present in the TS. This inference, however, should be viewed with caution, as the magnitude of the isotope effect is small (Fig. 3.11*d*, Table 3.2).

The apparent discrepancy in H-bond content between the mutational and isotope studies for CI2 and Protein G may provide an example where side-chain interactions, which can be tertiary in nature, do not exclusively reflect the backbone H-bond network, or that ϕ_f -values under-report structure (Bulaj & Goldenberg, 2001). Lower mutational ϕ_f -values in the helix alongside a higher isotope effect value might indicate that H-bond formation occurs with fewer native-like *side chain* contacts at the rate-limiting step. This uncertainty can be resolved using engineered metal binding sites and ψ -analysis, a methodology which can identify the fraction of pathways having the helix in a native conformation (Krantz & Sosnick, 2001).

3.7.10 *Protein L and SH3*

Unfortunately, the backbone amide isotope effects for *src*-SH3 and Protein L are too small and inconsistent to unambiguously interpret (Fig. 3.11*e,f*, Table 3.2). This is primarily due to the fact that there are few α -helical H-bonds and the more frequent β -sheet H-bonds lack a significant isotope effect. These proteins, however, were included to better assess the relationship of secondary structure composition to equilibrium isotope effects (see Fig 3.8).

3.7.11 *CD2 and hen egg white lysozyme*

Similar kinetic amide isotope measurements on CD2 (Parker & Clarke, 1997) and lysozyme (Itzhaki & Evans, 1996) did not to reveal any measurable difference in folding

rates between the deuterated and proteated forms. Potentially for CD2, this was due to a slight inequality in the bulk solvent conditions which obscured a presumably small isotope effect in this all β protein. The studies of lysozyme utilized identical bulk solvent conditions, but observed refolding rates that were indistinguishable (Itzhaki & Evans, 1996). While it is expected from present data that the purely β -sheet CD2 would not have an isotope effect, the results for the largely α -helical lysozyme are puzzling and may imply a general lack of experimental precision required for measuring such effects.

3.8 Transition state structure and general folding models

The kinetic amide isotope effect provides a direct, global probe for the extent of helical H-bond formation in the TS. We have now studied numerous two-state folding proteins with varying secondary structure compositions. At the rate-limiting step for largely α -helical proteins, λ -repressor, Arc repressor, and their variants as well as for IM7, Cyt *c* and GCN4, the degree of helix formation, given by $\phi_f^{\text{D-H}}$, is proportional to the degree of denaturant sensitive surface area buried, as quantified by m_f/m° (0.78 R-value, Fig. 3.12). In short, secondary structures and tertiary structures seem to form concomitantly in predominantly α -helical proteins.

However, for the β -sheet and mixed α/β systems, CI2, CTACP, Protein G, Protein L, SH3 and Ub, the correlation between $\phi_f^{\text{D-H}}$ and m_f/m° is poor (0.24 R-value). Proteins with extensive β -sheet networks would not be expected to exhibit such a correlation, because β -sheet does not have an appreciable isotope effect.

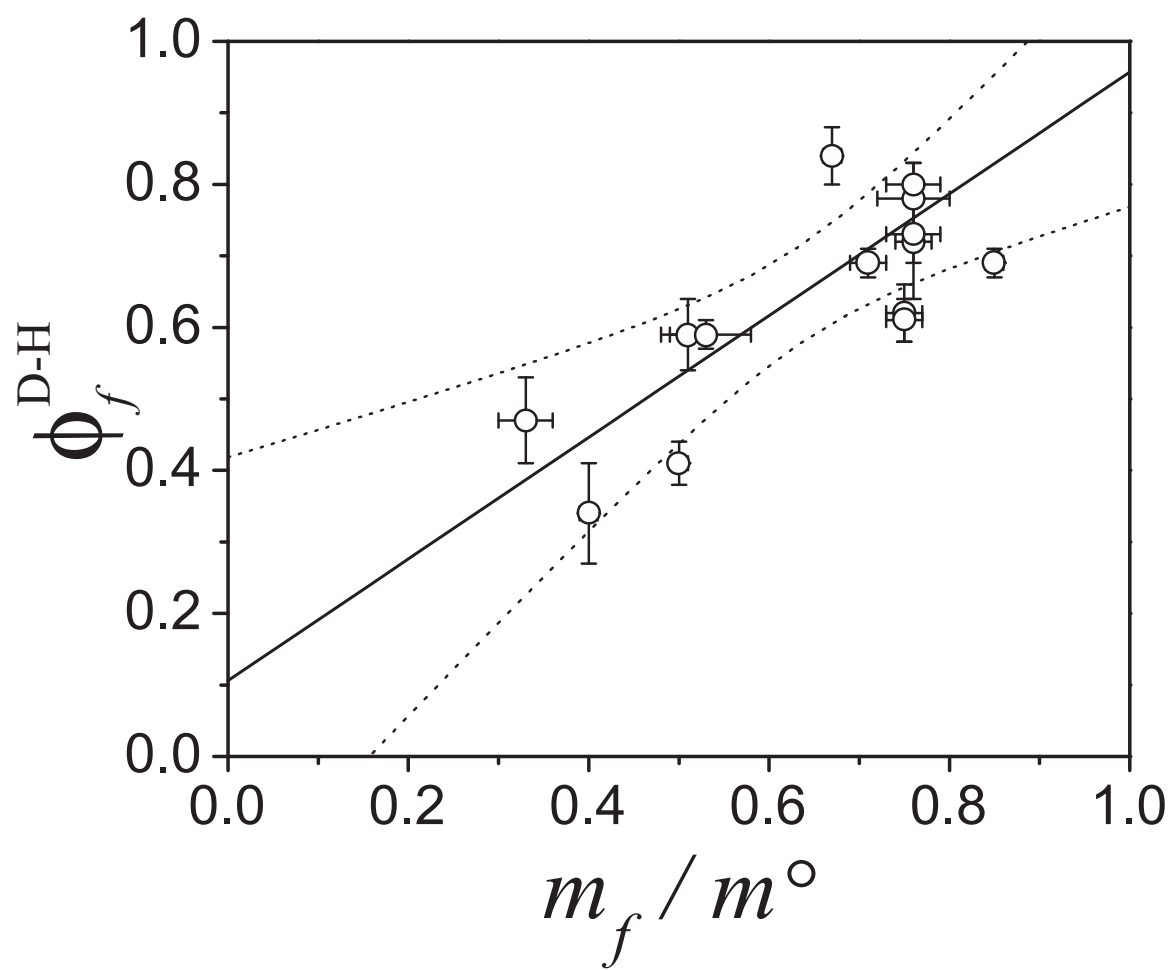
The correlation of $\phi_f^{\text{D-H}}$ and m_f/m^o values allows us to evaluate the applicability of various folding models for α -helical proteins. Had extensive hydrophobic collapse occurred prior to helix formation, the percentage of H-bonds formed should have been less than the amount of surface area burial. Furthermore, such a collapse even with marginal backbone desolvation is very unstable, as the failure to form each compensating intra-molecular H-bond costs several kcal mol⁻¹. Both of these observations argue against a non-specific hydrophobic collapse model.

Likewise, had extensive H-bond formation preceded tertiary interactions, the percentage of H-bond formed should have been larger than we observed. Also, folding contingent on pre-formed helical structure is contrary to results with Arc (Waldburger *et al.*, 1996), as discussed above, and studies of the dimeric GCN4 coiled coil (Moran *et al.*, 1999) and a variant with enhanced tertiary interactions (Durr *et al.*, 1999). Additionally, folding rates for Arc and GCN4, like most helical proteins, are faster at higher temperatures, contrary to the expectation for a simple folding scenario where rates are limited by pre-collision helical structure (i.e., as temperature increases, helical content decreases more significantly than the diffusion rates increases).

The interpretation that the TS is a critically distorted version of an essentially native-like structure with a desolvated core (e.g. the “dry molten-globule” model) also is not operational for helical proteins, because many H-bonds are not formed, and significant surface remains exposed in the TS.

Figure 3.12 H-bond formation and surface burial are concomitant folding processes.

Denaturant sensitive surface area burial (m_f/m^o) is plotted versus folding H-bond kinetic isotope effect ($\phi_f^{\text{D-H}}$) for largely α -helical proteins. A least squares linear fit (—) has a slope of 0.85 ± 0.19 , intercept of 0.11 ± 0.13 , and an R-value of 0.78, with 97% upper and lower confidence bands (\cdots).



The present data do support the hypothesis that protein folding is an incremental build-up of structure where hydrophobic burial and H-bond formation occur commensurately (Krantz *et al.*, 2000), rather than a model where extensive formation of either one of these quantities occurs prior to the other. The TS has extensive regions of presumably near-native structure, while other regions remain largely unfolded and exposed to solvent, as observed in unfolding intermediates of Cyt *c* (Bai *et al.*, 1995b) and ribonuclease A (Chamberlain *et al.*, 1996) identified using native state HX methods.

Previously, we proposed that the critical element of the TS is the formation of the over-all chain topology, which is established by pinning the chain by the interaction of a number of apolar side chains to adequately define the native topology (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995). At this nucleation event, a considerable amount of configuration entropy is lost and folding can proceed rapidly downhill. Consistently, a strong correlation was observed for proteins between the folding speed and the average sequence distance between residues, or contact order (Plaxco *et al.*, 1998). A large amount of other data also supports the importance of topology in folding (Chiti *et al.*, 1999; Guerois & Serrano, 2000; Martinez *et al.*, 1998; McCallister *et al.*, 2000; Moran *et al.*, 1999).

At the time of our original proposal (Sosnick *et al.*, 1996b; Sosnick *et al.*, 1995), we focused on TS topology driven by hydrophobic association but not by secondary structure formation. The present isotope study revises this view and establishes that hydrophobic association cannot be divorced from helix formation. Evidently, when the hydrophobic side chains of an amphipathic α -helix coalesce, water—the major competitors of backbone H-bonds—are expelled, which drives H-bonding between the amide and carbonyl groups. Once these H-bonds are formed, the solvent exposed H-

bonds on the opposite side of the now nucleated helices form as well. In this view, helix formation provides an efficient means to satisfy the H-bonding requirements of hydrophobically collapsed polypeptide. These principles should apply to H-bond formation in β -sheets as well, even though the edges can be solvated. We expect, therefore, that inter-strand H-bonds will be formed in the TS if they are heavily desolvated (see also Ch. 2).

Integration of these new results indicates that extensive backbone desolvation is a critical component of the TS, consistent with our conclusions regarding the effects of TFE on folding rates (Kentsis & Sosnick, 1998). Isolated helix formation does not drive the folding process. Rather, folding occurs as a result of hydrophobic association. The energetic benefit of hydrophobic association offsets the otherwise unfavorable formation of *isolated* helical structure. This situation is analogous to helix formation inside a membrane. Here, the equilibrium constant for helix formation is greatly increased as the alternative, exposed backbone groups, is now highly unfavorable (as opposed to the alternative view that native H-bonds become stronger in the lower dielectric environment). Thus, the reality that H-bonds form in proportion to surface area burial may simply reveal that the exclusion of water molecules necessitates the formation of H-bonds rather than the intrinsic strength of helix formation drives its own formation.

During the folding process, the polypeptide creates its own desolvated environment (Fernandez *et al.*, 2001). Transition states generally bury a significant amount of surface area, as indicated by the general observation that m_f -values are minimally about 1 and 0.5 units in GdmCl and urea, respectively. Hence, there may be a size threshold for backbone desolvation. Which regions become desolvated and result in structure formation in the TS, depends on the protein's sequence and structure. Without the formation of correct

topology to support productive and downhill folding for a given protein, apolar interactions will disassemble and re-assemble while searching for a collapse-competent nucleus. This topological requirement, which depends upon a protein's structure, may explain the wide diversity in the degree of denaturant sensitive surface area burial and extent of H-bond formation, from ~40% to ~85%. Likewise, a topological requirement rationalizes the observation that the amount of surface burial in the TS is typically independent of whether the protein is under stabilizing or destabilizing conditions.

3.9 Conclusions

Amide isotope effects provide a unique, powerful method to assess global H-bond formation in α -helices with minimal perturbation to the system. We find that helical H-bond formation and surface area burial occur in a concerted fashion. Models heavily weighting either of these two quantities appear contradictory to our results. Backbone desolvation accompanying hydrophobic association results in helix formation, as it is an efficient means to satisfy the H-bonding requirements. In combination with site-specific data, these insights, therefore, may best direct theoretical simulations.

3.10 Methods

3.10.1 *Proteins*

Protein expression plasmids were constructed using pRSET B (Invitrogen) into which open reading frames for IM7, λ^{AA} , CI2, CTACP, and SH3 proteins were subcloned. Open reading frames encoding protein sequences were generated using total DNA synthesis (Operon technologies, see also App. 1). Protein sequences (Table A2.2 in

App. 2) are described in previous publications or references therein (Burton *et al.*, 1997; Ferguson *et al.*, 1999; Grantcharova & Baker, 1997; Itzhaki *et al.*, 1995b; Taddei *et al.*, 2000). Ubiquitin, Ub F45W, (Krantz & Sosnick, 2000), Protein L, Protein G, Arc and Arc^{MYL} were obtained as previously described (Nauli *et al.*, 2001; Waldburger *et al.*, 1996). A λ -repressor (6-85) 'wild-type' version, λ^{GG} , was created using the Stratagene Quick Change mutagenesis kit to create the double A46G A48G substitutions. Protein purification from standard *E. coli* expression lysates were carried out using a combination of anion or cation exchange chromatography (Pharmacia) followed by C18 reverse phase HPLC (Vydac) in acetonitrile with 0.1% trifluoroacetic acid. Nickel affinity chromatography (Pharmacia) purified Protein L and Protein G, which possess six histidine N-terminal tags. GCN4^{Cx}_{biH3,7} is a C-terminally disulfide bond cross-linked version of GCN4-p1' described previously (Krantz & Sosnick, 2001). All proteins were lyophilized and analyzed by mass spectrometry. Deuterated versions of proteins were created by dissolving lyophilized protonated protein in deuterated 7.4 M GdmCl buffer, allowing a minimum of one hour for complete exchange at room temperature.

3.10.2 Equilibrium isotope effect measurements

CD monitored change in protein equilibrium stability during HX, at wavelengths ranging from 222 to 228 nm (2 to 5 nm bandwidth) using a Jasco J-715 spectropolarimeter. Peptide concentrations ranged from 2 to 10 μ M.

3.10.3 Stopped-flow kinetics

Rapid mixing experiments used a Biologic SFM4 stopped-flow apparatus (Krantz & Sosnick, 2000). Protein concentrations ranged from 0.1 μ M to 10 μ M.

The data were analyzed using chevron analysis of the denaturant dependence of folding rate constants (Matthews, 1987) where the standard free energy of folding, ΔG^o , along with the standard activation free energy for folding ΔG_f^\ddagger , and unfolding, ΔG_u^\ddagger , are linearly dependent on denaturant concentration (see App. 3). When equilibrium and kinetic folding reactions are effectively two-state and are limited by the same activation barrier, the equilibrium values for the standard free energy and surface burial can be calculated from kinetic measurements according to $\Delta G^o = \Delta G_f^\ddagger - \Delta G_u^\ddagger$ and $m^o = m_u - m_f$. Parameters were fit using non-linear least-squares algorithms implemented in the Microcal Origin software package.

4

Engineered Divalent Metal Ion Binding Sites and Pathway Heterogeneity

4.1 Abstract

To address whether proteins fold along multiple pathways, i,i+4 bi-histidine metal binding sites are introduced in dimeric and cross-linked versions of the leucine zipper region of transcription factor GCN4 and the mixed α/β protein, ubiquitin. Divalent metal ion binding enhances both the equilibrium and folding activation free energies. The enhancement of folding rates quantifies the fraction of molecules that have the binding site in a helical geometry in the transition state. Hence, this new method, termed ψ -analysis, identifies the degree of pathway heterogeneity for a protein that folds in a two-state manner, a capability that is generally unavailable even with single-molecule methods. Adjusting metal ion concentration continuously varies the stability of the bi-histidine region without additional structural perturbation to the protein. For GCN4 and ubiquitin, the accompanying changes in kinetic barrier heights at each metal ion concentration maps the folding landscape as well as establishes the importance of connectivity in pathway selection. Furthermore, this method can be generalized to other biophysical studies, where the ability to continuously tune the stability of a particular region with no extraneous structural perturbation is advantageous.

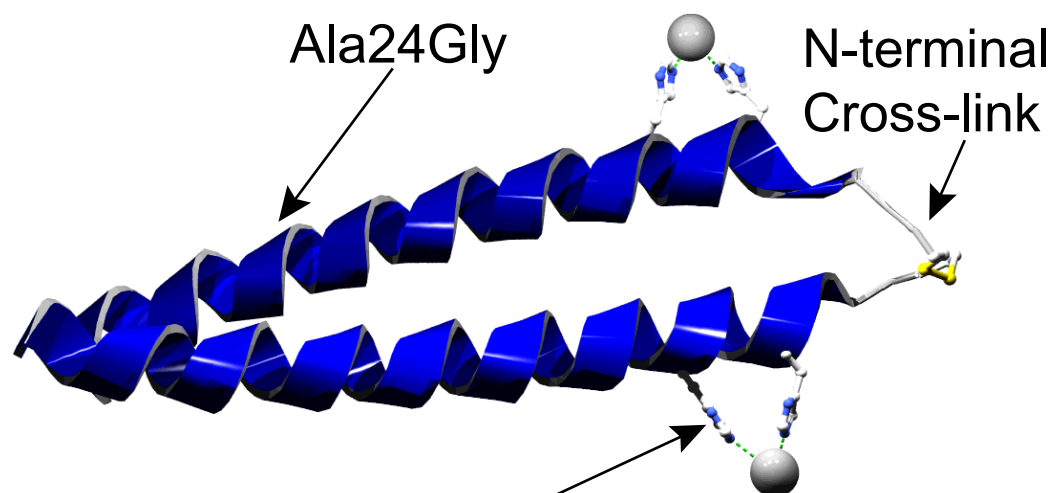
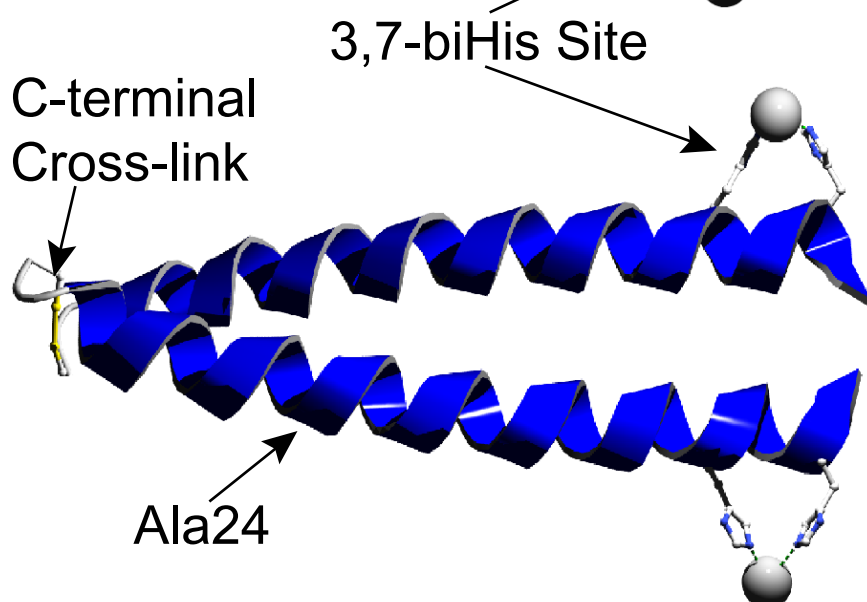
4.2 Introduction

Traditionally, protein folding is considered a determinate process, whereby specific intermediates populate along a singular pathway. Most (Fersht *et al.*, 1994; Grantcharova *et al.*, 2000; Martinez *et al.*, 1998), but not all (Burton *et al.*, 1998; Moran *et al.*, 1999) folding experiments are interpreted in the context of a homogeneous transition state (TS) ensemble. Theoretical work, however, has led to a funnel picture in which folding occurs via structurally distinct, heterogeneous routes (Dill & Chan, 1997; Shakhnovich, 1998; Socci *et al.*, 1998; Thirumalai & Klimov, 1998). This controversy, termed The Classical versus New View Debate, has become one of the most discussed in protein folding (Baldwin, 1995; Dill & Chan, 1997).

Debate has continued, because it is experimentally difficult to resolve this issue for two-state folding proteins. For example, mutational ϕ -analysis, the most frequently applied method to characterize TS's, is confounded by multiple interpretations of intermediate ϕ_f -values. The energetic effect of an amino acid substitution on the folding activation energy relative to equilibrium stability, quantified as ϕ_f , defines the level of the mutated residue's interaction in the TS. A ϕ_f -value of zero or one indicates that the side-chain's influence is either absent or fully realized, respectively. Fundamental ambiguities, however, arise for the fractional values normally observed. Do these values represent a single TS with partially formed structure or a population of heterogeneous TS's?

We resolved this apparent ambiguity for the tryptophan-containing α -helical coiled coil GCN4-p1' (Sosnick *et al.*, 1996a), using single and double Ala to Gly substitutions. We deduced that fractional values in the dimeric protein were due to multiple pathways

Figure 4.1 Designed metal ion binding variants of the coiled coil. Depictions of cross-linked coiled coil with the 3,7-biHis metal binding site and either an N-terminal CGG disulfide tether containing the A24G mutation **(a)** or a C-terminal GGC disulfide tether **(b)**. Dimeric versions of either cross-linked species are identical peptides except each lacks the terminal cysteine residue required to make the cross-link. Graphics were rendered by Swiss-Prot Protein Viewer (Glaxo Wellcome Experimental Research) and the Persistence of Vision ray trace program (POV-Ray).

a*b*

with helix nucleation sites located along the length of the coil. After inserting a destabilizing mutation at one site, which possessed an intermediate ϕ_f -value and was potentially important in the TS, we conducted ϕ -analysis at a second site, which previously had a small ϕ_f -value. The ϕ_f -value at the second site subsequently increased significantly, indicating that helix nucleation had shifted toward this region.

Given the importance of identifying pathway heterogeneity and the ubiquity of fractional ϕ_f -values, we developed a simple method that characterizes the topography of the folding landscape. Divalent metal ion binding sites are engineered at solvent exposed positions to stabilize specific regions (Fig. 4.1). The relative enhancement of folding rates with respect to increasing binding energy identifies the percentage of TS's with the metal binding site formed, similar to ligand binding studies at naturally-occurring sites (Sancho *et al.*, 1991; Segawa & Sugihara, 1984). The binding energy required to switch a low flux pathway to a high flux pathway defines the relative barrier heights between the pathway with the metal site formed relative to all other pathways. Hence, within the context of a single protein, this method can distinguish single pathway models with partial structure formation in the TS from multiple pathway folding models as well as map the free energy reaction surface.

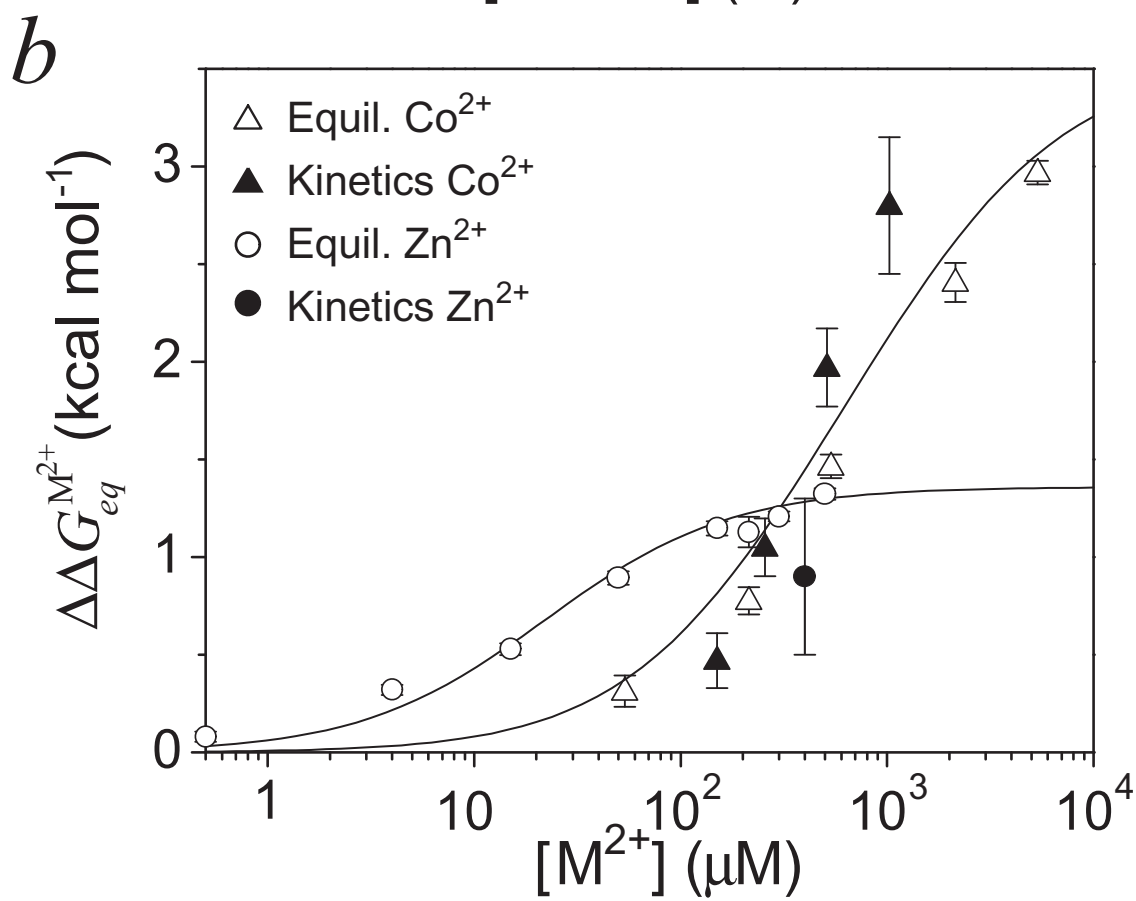
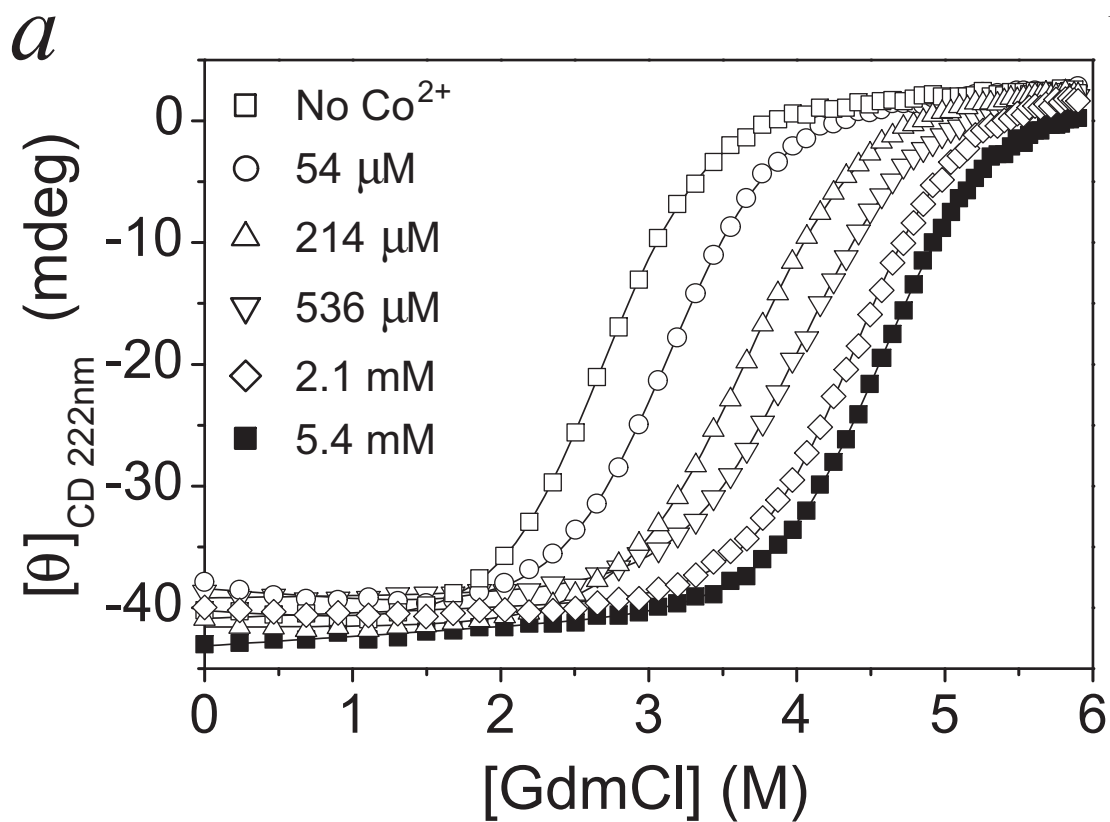
4.3 The coiled coil, GCN4

4.3.1 Equilibrium binding stability of designed binding sites

We introduced a pair of bi-histidine (biHis) metal binding sites in an $i,i+4$ configuration (Ghadiri & Choi, 1990) into GCN4-p1' (Fig. 4.1). This design is borrowed from zinc finger motifs possessing two histidines on a helix, which coordinate the Zn^{2+}

Figure 4.2 Equilibrium stability obtained by the addition of the divalent metal ions. (a)

Standard GdmCl denaturation melts of GCN4^{Nx,A24G}_{biH3,7} in either no Co²⁺ (□), 54 μM Co²⁺ (○), 214 μM Co²⁺ (△), 536 μM Co²⁺ (▽), 2.1 mM Co²⁺ (◇), or 5.4 mM Co²⁺ (■). The ~4% baseline variance in CD level at low denaturant is nonsystematic, and presumably due to a slight variation in protein concentration, rather than a structural perturbation. This interpretation is supported by a measurement on the dimeric A24 version (data not shown) where any potential effect of the N-terminal tether is removed. The addition of metal stabilizes the dimeric version by ~ 2.8 kcal mol⁻¹, which increases the folded population by 4%, and accounts for nearly all of the 5-6% change in CD signal. *(b)* Fits to standard denaturant melts above (*solid lines*) yield the equilibrium stability for each concentration of Co²⁺ (△) and Zn²⁺ (○) for GCN4^{Nx,A24G}_{biH3,7}. For comparison, equilibrium stability values estimated kinetically are plotted as well for Co²⁺ (▲) and Zn²⁺ (●). The data for each metal ion are fit to Eq. 4.1, and fit parameters for all metal ion binding versions of GCN4 studied are contained in Table 4.1. Note that while Zn²⁺ binds more tightly to the biHis site it yields less stabilization energy than Co²⁺ when saturated, since the overall gain in free energy is due to that ratio of the K_d for the folded and unfolded states.



ion along with two cysteine residues on a β -turn. In the presence of Co^{2+} and Zn^{2+} at pH 7.5, where histidines are de-protonated, metal binding stabilizes the coiled coil up to ~ 3.6 kcal mol $^{-1}$ (Fig. 4.2b, Table 4.1). The degree of stabilization for the two independent sites, $\Delta\Delta G_{eq}$, depends upon differences in metal dissociation constants between the two biHis sites in the native state, K_{eq}^N , and the two sites in the denatured state, K_{eq}^U .

$$\Delta\Delta G_{eq}(\text{M}^{2+}) = 2RT \ln(1 + [\text{M}^{2+}] / K_{eq}^N) - 2RT \ln(1 + [\text{M}^{2+}] / K_{eq}^U) \quad (4.1)$$

For Co^{2+} , binding constants in the native and denatured states are in the 50 μM and 1 mM ranges, respectively. While binding is five to ten-fold stronger for Zn^{2+} , the ratio of binding in the native state to the denatured state is less, producing less added stability.

For the N-terminally cross-linked A24G form, circular dichroism (CD) signal at 222 nm (helical structure) for the native state is insensitive to metal concentration (Fig. 4.2a). Although the baselines in the denaturation profiles have a standard deviation of 4%, there is no systematic relation to metal concentration, indicating that metal binding to the biHis site does not introduce any appreciable structural perturbation. The CD signal for the denatured state is unchanged at different cation concentrations, indicating that binding to the unfolded state does not occur in a helical geometry. Addition of 1 mM Co^{2+} to wild-type GCN4-p1', which contains a single solvent accessible histidine on each helix, fails to alter stability, indicating that metal-induced stabilization is specific to i,i+4 biHis substitution. This result suggests that non-neighboring surface histidine residues can be retained for these studies. However, we introduced a H18N substitution to completely preclude any unwanted complications. We presume that the analysis will be more complicated for proteins that naturally bind metals.

Table 4.1 Equilibrium metal ion stabilization of biHis GCN4 variants^a

Protein ^b	Divalent metal ion (M ²⁺)	K_{eq}^N (μM)	K_{eq}^U (μM)
GCN4 ^{Nx,A24G} _{biH3,7}	Co ²⁺	129±29	2980±1800
GCN4 ^{Nx,A24G} _{biH3,7}	Zn ²⁺	11.5±1.7	37.1±6.3
GCN4 ^{Cx} _{biH3,7}	Co ²⁺	44.4±7.8	418±155
GCN4 ^{d,A24G} _{biH3,7}	Co ²⁺	65.8±5.6 ^c	1660±319 ^c
GCN4 ^d _{biH3,7}	Zn ²⁺	4.90±0.14 ^d	36.5±1.3 ^d

^aEquilibrium metal ion binding constants are obtained from data similar to that shown in Fig. 4.2, fit according to Eq. 4.1, which assumes a single metal ion is bound at each pair of histidines in the native state, K_{eq}^N , and a ion is bound at each pair of histidines in the unfolded state, K_{eq}^U . Other binding models could fit the data adequately (e.g. one ion per histidine in the unfolded state), but none of the conclusions depends on the functional form, as only the net added stability, $\Delta\Delta G_{eq}^{M^{2+}}$, is used in the analysis.

^bGCN4 topologies are depicted in Fig. 4.1, and protein sequences are given in App. 2.

^cGCN4^{d,A24G}_{biH3,7} experiments are conducted at 40 μM monomer concentration.

^dGCN4^d_{biH3,7} experiments are conducted at 10 μM monomer concentration.

Figure 4.3 Denaturant dependence of folding in presence of metal ions for cross-linked

GCN4. (a) Kinetic activation energy plots, or chevrons, for GCN4^{Nx,A24G}_{biH3,7} with either no

Co²⁺ (□), 150 μM Co²⁺ (○), 256 μM Co²⁺ (△), 512 μM Co²⁺ (▽), or 1 mM Co²⁺ (◇).

(b) Kinetic activation energy plots, or chevrons, for GCN4^{Cx}_{biH3,7} with either no metal

(□), 100 μM Zn²⁺ (○), 400 μM Zn²⁺ (△), or 1 mM Co²⁺ (▽). Data are fit to the two-

state chevron model (App. 3) and curve fit parameters are listed in Table 4.2.

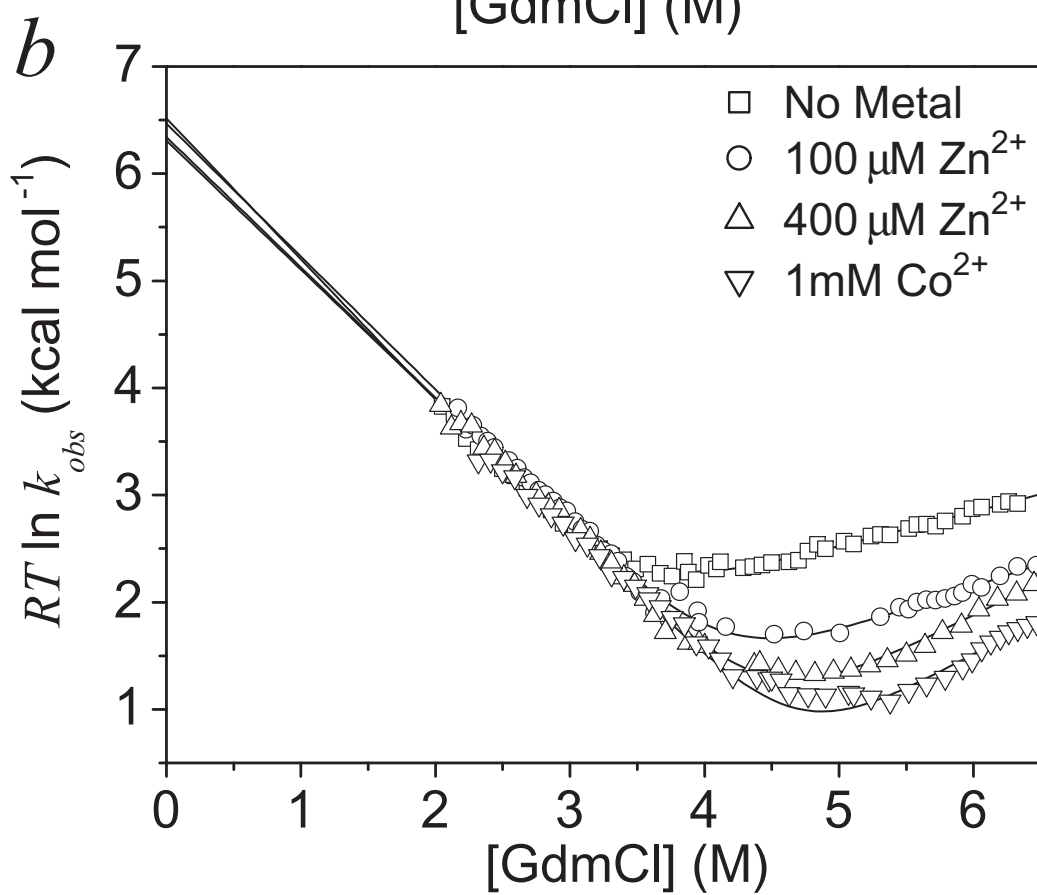
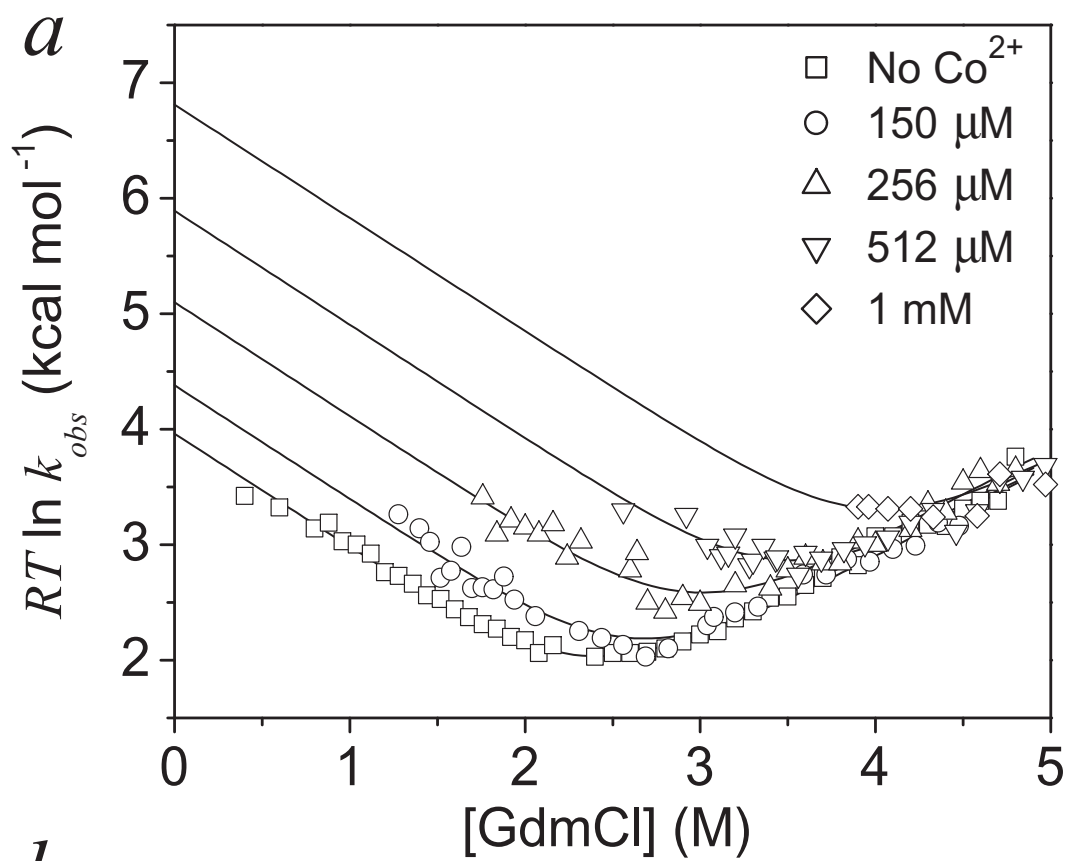


Table 4.2 Kinetic parameters for metal ion binding to biHis GCN4 variants

Protein ^a	[M ²⁺] (μ M)	$RT \ln k_f$ (kcal mol ⁻¹)	ΔG^o (kcal mol ⁻¹)	$-m_f$ (kcal mol ⁻¹ M ⁻¹)	m^o (kcal mol ⁻¹ M ⁻¹)
GCN4 Nx,A24G biH3,7	No metal	3.96±0.06	-4.17±0.10	0.98±0.04 ^e	1.77±0.04 ^e
	150 Co ²⁺	4.64±0.08	-4.64±0.10	0.98±0.04 ^e	1.77±0.04 ^e
	256 Co ²⁺	5.10±0.09	-5.22±0.11	0.98±0.04 ^e	1.77±0.04 ^e
	512 Co ²⁺	5.89±0.13	-6.14±0.13	0.98±0.04 ^e	1.77±0.04 ^e
	1024 Co ²⁺	6.81±0.17	-6.96±0.17	0.98±0.04 ^e	1.77±0.04 ^e
GCN4 ^{Cx} biH3,7	No metal	6.52±0.08	-5.63±0.07	1.32±0.03 ^f	1.64±0.03 ^f
	100 Zn ²⁺	6.47±0.08	-6.89±0.23	1.24±0.03 ^f	1.67±0.04 ^f
	400 Zn ²⁺	6.31±0.06	-8.19±0.13	1.20±0.02 ^f	1.83±0.03 ^f
	1000 Co ²⁺	6.34±0.06	-9.01±0.11	1.22±0.02 ^f	1.91±0.03 ^f
GCN4 ^{d,A24G} biH3,7	No metal	0.76±0.04 ^{b,c}	-1.35±0.04 ^{b,c}	N.D. ^g	N.D.
	80 Co ²⁺	1.34±0.03 ^{b,c}	-2.19±0.02 ^{b,c}	N.D.	N.D.
	160 Co ²⁺	1.71±0.04 ^{b,c}	-2.64±0.03 ^{b,c}	N.D.	N.D.
	250 Co ²⁺	1.99±0.03 ^{b,c}	-2.96±0.03 ^{b,c}	N.D.	N.D.
	500 Co ²⁺	1.95±0.03 ^c	-2.91±0.05 ^c	0.67±0.03 ^e	1.58±0.04 ^e
	1500 Co ²⁺	2.18±0.04 ^c	-3.14±0.05 ^c	0.67±0.03 ^e	1.58±0.04 ^e
	10 ⁴ Co ²⁺	2.62±0.03 ^c	-3.58±0.05 ^c	0.67±0.03 ^e	1.58±0.04 ^e
GCN4 ^d biH3,7	No metal	0.95±0.08 ^d	-1.50±0.09 ^d	1.09±0.11 ^f	1.70±0.11 ^f
	25 Co ²⁺	0.96±0.06 ^d	-1.98±0.06 ^d	1.02±0.14 ^f	1.55±0.14 ^f
	50 Co ²⁺	1.04±0.03 ^d	-2.29±0.06 ^d	1.19±0.06 ^f	1.72±0.07 ^f
	100 Co ²⁺	1.05±0.02 ^d	-2.46±0.06 ^d	1.15±0.04 ^f	1.65±0.05 ^f
	250 Co ²⁺	1.08±0.01 ^d	-3.09±0.05 ^d	1.30±0.03 ^f	1.91±0.04 ^f
	500 Co ²⁺	1.10±0.02 ^d	-3.52±0.06 ^d	1.27±0.04 ^f	1.97±0.04 ^f
	1000 Co ²⁺	1.20±0.01 ^d	-3.82±0.07 ^d	1.22±0.02 ^f	1.95±0.03 ^f
	2000 Co ²⁺	1.34±0.01 ^d	-4.05±0.11 ^d	1.21±0.02 ^f	1.97±0.07 ^f
	4000 Co ²⁺	1.48±0.01 ^d	-4.24±0.24 ^d	1.23±0.03 ^f	1.98±0.07 ^f
	7500 Co ²⁺	1.55±0.01 ^d	-4.31±0.15 ^d	1.22±0.04 ^f	1.98±0.06 ^f

^aGCN4 topologies are depicted in Fig. 4.1, and protein sequences are given in App. 2.

^bEstimated GCN4^{d,A24G}_{biH3,7} folding activation energy, $RT \ln k_f$ from equilibrium melting data and unfolding activation energy.

^cGCN4^{d,A24G}_{biH3,7} refolding activation energy and equilibrium stability values for the experimental condition, 31.7 μ M final monomer concentration.

^dGCN4^d_{biH3,7} refolding activation energies and equilibrium stability values for the experimental condition, 2 μ M final monomer concentration.

^eGlobal fits for all metal ion conditions produced the best fit m_f and m^o values.

^fIndependently fit m_f and m^o values for each metal ion condition.

^gNot determined.

4.3.2 Cross-linked coiled coil and pathway homogeneity

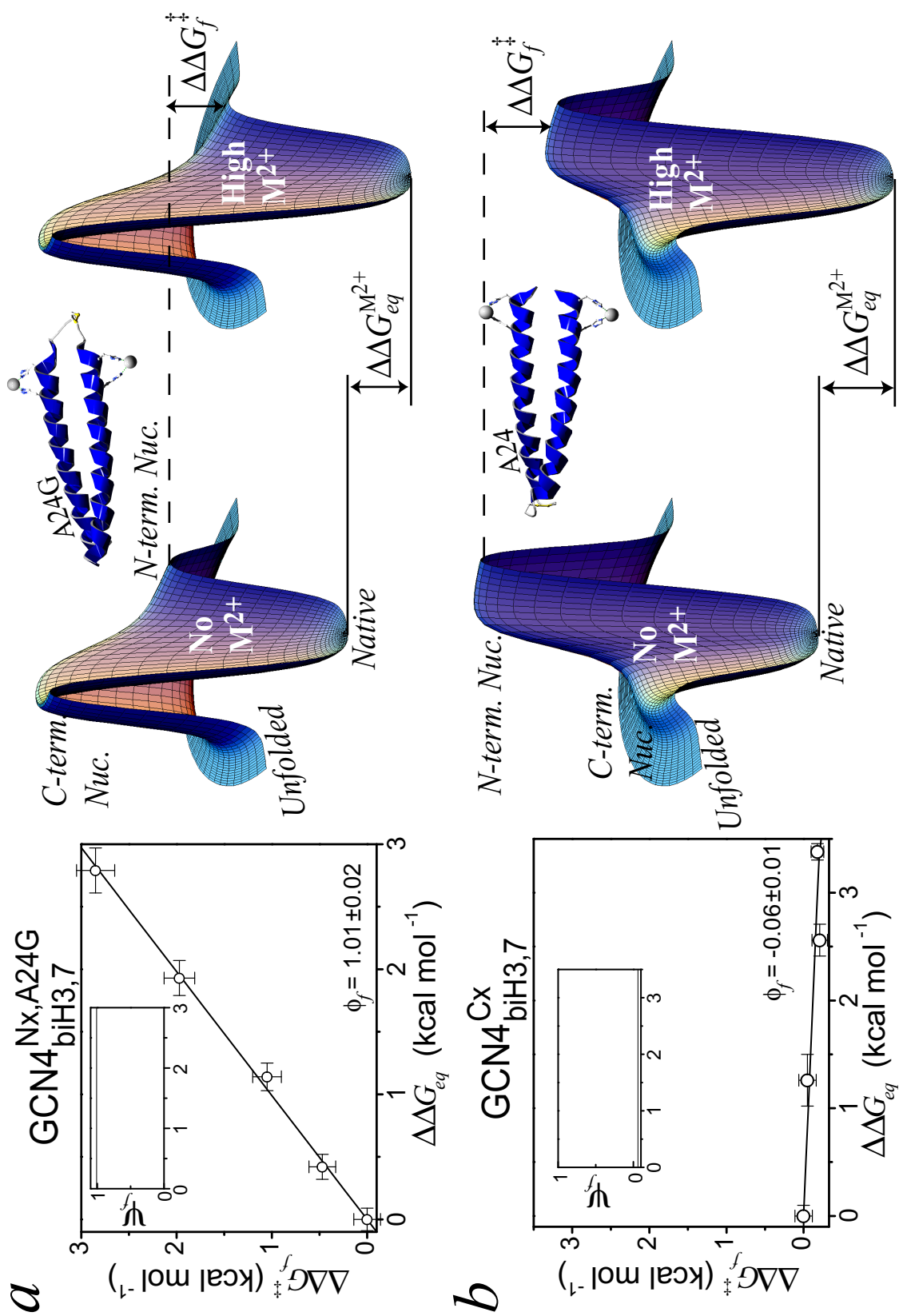
BiHis metal-binding studies are analogous to mutational ϕ -analysis, but, instead of an amino acid substitution, divalent metal ion concentration is varied to perturb the stability of a specific region. The degree to which the region is structured in the TS at a given metal concentration is the relative effect on folding activation free energy versus the change in equilibrium stability, $\phi_f^{M^{2+}} = \Delta\Delta G_f^\ddagger / \Delta\Delta G_{eq}$. We first illustrate this method with the N-terminal cross-linked version (Fig. 4.1a), as its folding behavior is straightforward, occurring along a single robust pathway, unlike that for the dimeric version which exhibits more complex, multiple route behavior (Moran *et al.*, 1999). According to our mutational studies on the N-terminal cross-linked version, the region nearest the cross-link is structured in the TS, while the region farthest from the tether is unstructured.

For biHis metal binding sites located near the amino-tethered end (residues 3 and 7), the addition of metal only accelerates folding rates, accounting for the entire change in equilibrium stability. In the denaturant dependence of activation energy plots, the folding arm increases with increasing metal concentration while the unfolding arms remain unchanged (Fig. 4.3a). The $\phi_f^{M^{2+}}$ -value is unity at all metal concentrations (Fig. 4.4a). Hence, the folding pathway is robust and homogeneous, where the entire TS ensemble possesses helix at the 3,7-biHis region (Fig. 4.4a).

To investigate the influence of connectivity, the cross-link is moved to the C-terminus, leaving the 3,7-biHis site intact (Fig. 4.1b). For this version, the $\phi_f^{M^{2+}}$ -value is zero in saturating concentrations of Zn^{2+} or Co^{2+} (Fig. 4.3b). This result indicates that folding remains homogeneous, and helix nucleates only at the tethered end, far from the

Figure 4.4 Determining folding pathway models for cross-linked GCN4. Change in

folding rates versus changes in stability due to metal binding (*left*) for GCN4^{Nx,A24G}_{biH3,7} **(a)** and **(b)** GCN4^{Cx}_{biH3,7}. Linear fits (**—**) are for an ideal homogeneous folding pathway, which best represents the data for cross-linked GCN4. *Insets*: Relative flux, ψ_f , through the metal-bound pathway is plotted as a function of the change in stability, or $\Delta\Delta G_{eq}$ (Eq. 4.5). Because cross-linked versions have robust homogeneous pathways, their respective ψ_f -values are constant and equal to ϕ_f -values. Folding energy landscapes (*right*) are modeled using ρ and ψ_f -values with two pathways, the N-terminal helix nucleation route containing the biHis site and the C-terminal helix nucleation route. The unfolded ensemble occupies the outer ring. When GCN4-p1' is cross-linked at the N-terminus **(a)**, the surface possesses a smaller barrier in the absence of M^{2+} for the N-terminal route with respect to the C-terminal route. Upon the addition of M^{2+} , the protein becomes more stable, and the N-terminal route's barrier is lowered and all the pathways continue to nucleate at the N-terminus. When M^{2+} is added to the C-terminal cross-linked version **(b)**, all the flux nucleates at the C-terminus, because the N-terminal barrier is too large even following 3 kcal mol⁻¹ stabilization. Graphics were created in Mathematica 4.0 (Wolfram Research, Inc.) using a cylindrical function that radially varies TS barrier height free energies in the z-axis by a smooth sigmoid relationship, defining the N- and C-terminal TS's.



biHis site (Fig. 4.4*b*). This shift is expected, because the metal site is now at the opposite end and distant from the cross-link. Hence, for the cross-linked coiled coil, folding is homogeneous with pathway selection dominated by chain connectivity (Fig. 4.4).

4.3.3 Quantifying heterogeneity: ψ -analysis

Although the interpretations are straightforward for the cross-linked versions' $\phi_f^{M^{2+}}$ -values of zero and unity, the folding of dimeric GCN4 is complex (Fig. 4.5). We present a complete derivation, similar to Brønsted analysis (Fersht *et al.*, 1994), for heterogeneous scenarios with fractional ϕ_f -values. Folding rates are calculated assuming two pathways, one where the biHis site is formed (k_1) and a second, representing the sum of all other pathways (k_2). The observed folding rate is

$$k_{obs} = k_1 + k_2. \quad (4.2)$$

In the absence of added metal, the relative flux is defined as $\rho = k_2/k_1$. Stabilization of the first pathway's TS due to metal binding, $\Delta\Delta G_f^{M^{2+}}$, enhances k_1 , and increases the overall rate according to

$$k_{obs}^{M^{2+}} = k_1 \left(\rho + \exp \left(f \times \Delta\Delta G_{eq} / RT \right) \right). \quad (4.3)$$

A term, f , is introduced to account for the possibility of fractional stabilization by the metal ion binding site in the TS, $\Delta\Delta G_f^{M^{2+}} = f \times \Delta\Delta G_{eq}$. Having ρ and f parameters resolves the inherent ambiguity in ϕ -analysis: fractional ϕ -values can result either from multiple pathways (ρ) and/or partial structure formation (f) in the TS. Empirically, for the solvent

exposed biHis site in the N-terminal cross-linked version (Fig. 4.1), the slope is unity, and therefore, f is unity (Fig. 4.4a). f -values, nonetheless, are provided in Table 4.3.

Upon addition of metal, the net reduction of the folding activation energy is

$$\Delta\Delta G_f^\ddagger = -RT \ln \left[\frac{\rho + 1}{\rho + \exp(f \times \Delta\Delta G_{eq} / RT)} \right]. \quad (4.4)$$

For standard mutational ϕ -analysis, the change in stability is a discrete value for a single mutation, because it is derived from a single comparison to wild-type protein. With biHis sites, analogous $\phi_f^{M^{2+}}$ -values can be calculated at each metal concentration, suitably referenced to the metal-free condition.

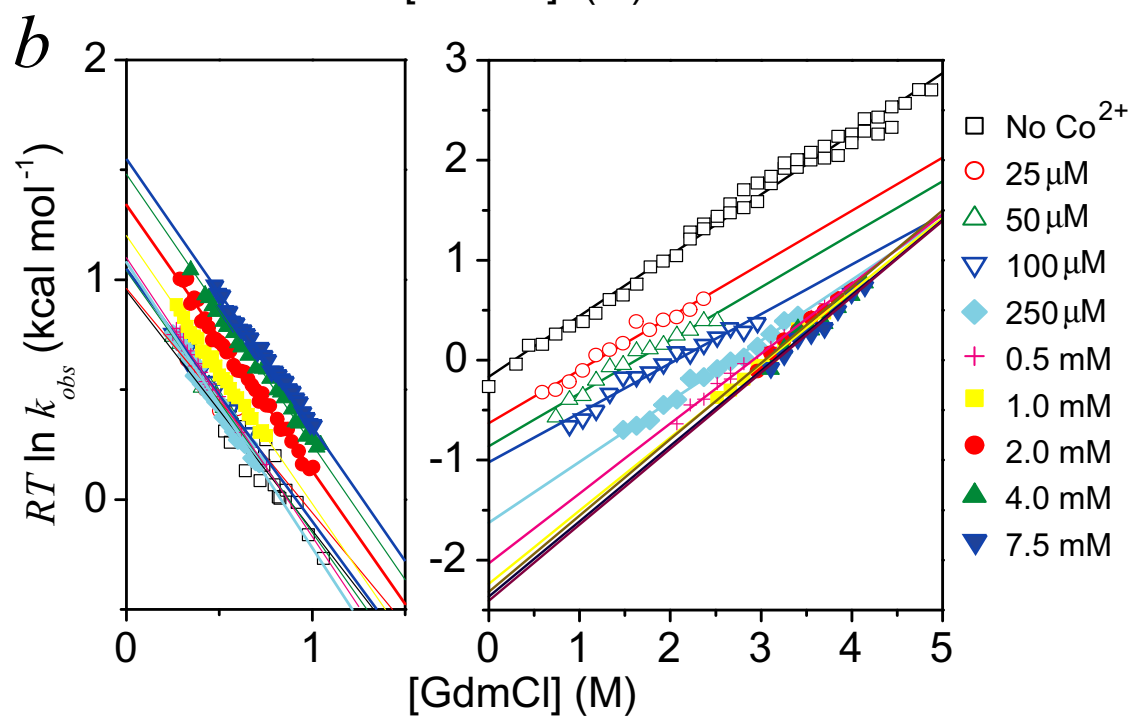
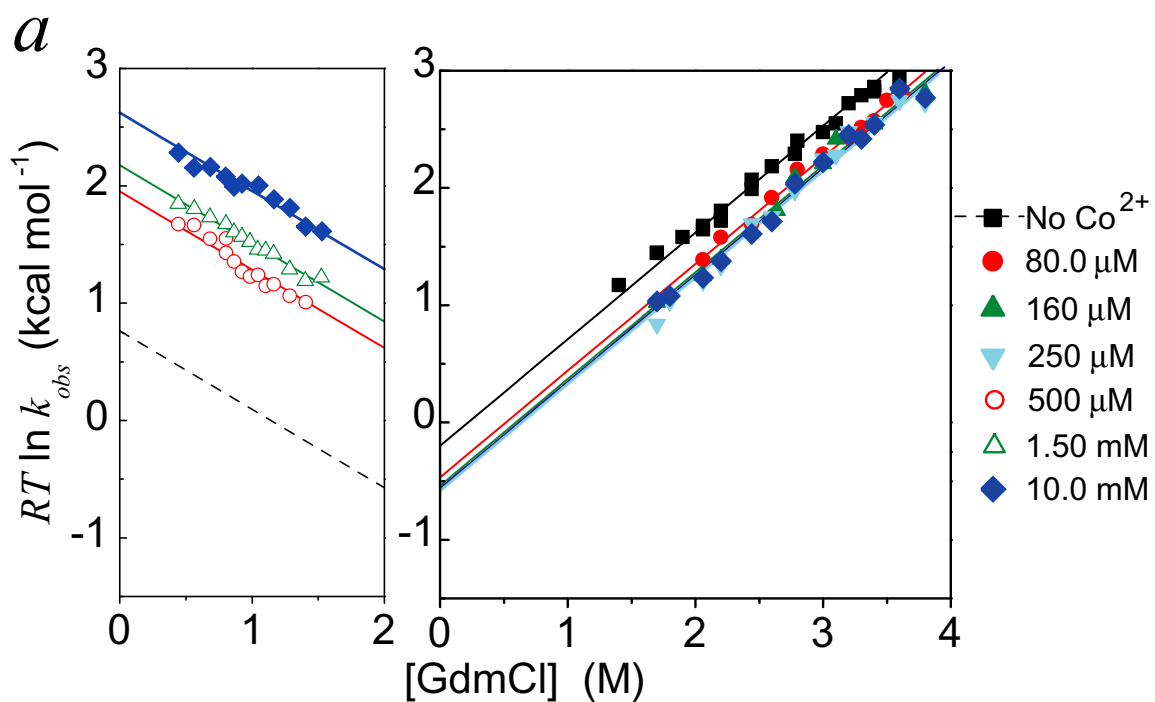
However, a much more illuminating quantity, ψ_f , may be calculated for each *incremental* change in metal ion concentration. This differential quantity is the infinitesimal change in activation energy relative to the infinitesimal change in equilibrium stability

$$\psi_f = \frac{\partial \Delta\Delta G_f^\ddagger}{\partial \Delta\Delta G_{eq}} = \frac{f \times \exp(f \times \Delta\Delta G_{eq} / RT)}{\exp(f \times \Delta\Delta G_{eq} / RT) + \rho}. \quad (4.5)$$

ψ_f is tangent to the curve of $\Delta\Delta G_f^\ddagger$ versus $\Delta\Delta G_{eq}$ (Fig. 4.6). In the simplified case of a homogeneous and robust folding pathway (i.e. ρ is either large, $\gg 100$, or small, $\ll 0.1$), ψ_f is either a constant of zero or f , depending on whether none of the pathways or all of the pathways have metal binding sites formed in the TS. For example, ψ_f for the N-terminal cross-linked version is unity as the N-terminal biHis region is always structured (Fig. 4.4a). Conversely, ψ_f is zero for the C-terminal cross-linked version as the N-terminal biHis region is completely unstructured at the rate-limiting step (Fig. 4.4b).

Figure 4.5 Denaturant dependence of folding in presence of metal ions for dimeric

GCN4. **(a)** Kinetic activation energy plots for GCN4^{d,A24G}_{biH3,7} for folding (*left*) and unfolding (*right*) with either no Co²⁺ (■), 80 μM Co²⁺ (●), 160 μM Co²⁺ (▲), 250 μM Co²⁺ (▼), 500 μM Co²⁺ (○), 1.5 mM Co²⁺ (△), or 1 mM Co²⁺ (◆). A reference line (--) for the folding denaturant dependence is provided and is estimated from the unfolding activation energy and the equilibrium stability (App. 3). **(b)** Kinetic activation energy plots, or chevrons, for GCN4^d_{biH3,7} for folding (*left*) and unfolding (*right*) with either no Co²⁺ (□), 25 μM Co²⁺ (○), 50 μM Co²⁺ (△), 100 μM Co²⁺ (▽), 250 μM Co²⁺ (◆), 0.5 mM Co²⁺ (+), 1.0 mM Co²⁺ (■), 2.0 mM Co²⁺ (●), 4.0 mM Co²⁺ (▲), or 7.5 mM Co²⁺ (▼). Data are fit to linear functions, using the two-state formalism (App. 3) and curve fit parameters are placed in Table 4.2.



4.3.4 Pathway heterogeneity in dimeric GCN4

Pathway homogeneity observed for cross-linked coiled coil is in contrast with pathway heterogeneity observed for dimeric GCN4 (Moran *et al.*, 1999), thereby providing a test of the method's ability to identify and quantify pathway heterogeneity. Initially, low concentrations of metal have a negligible impact on folding rates of 3,7-biHis dimeric GCN4 (Fig. 4.5b), and ψ_f remains near zero (Fig. 4.6b). Generally, ψ_f represents the ratio of the molecules that nucleate at the biHis region relative to those that nucleate at all other regions. Hence, the biHis route is a very minor pathway in the absence of metal. As metal concentration is increased, the N-terminal region is stabilized. More molecules nucleate at the N-terminal end, and ψ_f increases to about one-half at high metal concentrations (Fig. 4.6b). Although ψ_f should approach unity with further stabilization of the binding site, it does not because of the limited experimental range for $\Delta\Delta G_{eq}$. A detailed analysis of the metal dependence indicates that the biHis-containing pathway is 80-fold less populated than all other routes in the absence of metal (Table 4.3; Fig. 4.6b).

The rationale for the relative fluxes observed in the present work, $\rho_{\text{biHis}} \sim 80:1$, versus the previous study, $\rho_{\text{wild-type}} \sim 6:1$ estimated for the wild-type protein, is due to the introduction of the biHis site. The two histidines reduce the intrinsic helicity of the region by > 10 -fold. Thus, in the biHis version, some of the flux normally at the N-terminus has shifted back towards the C-terminus. This shift explains the reduction in the observed pathway heterogeneity in the absence of metal for the biHis protein.

Figure 4.6 Determining folding pathway models for dimeric GCN4. Change in folding rates versus changes in stability due to metal binding (*left*) for GCN4^{d,A24G}_{biH3,7} (**a**), and GCN4^d_{biH3,7} (**b**). Linear fits (--) are for a homogeneous pathway. Hyperbolic fits (—) represent a heterogeneous model for dimeric GCN4 (Eq. 4.4). *Insets:* Relative flux, ψ_f , through the metal-bound pathway is plotted as a function of the change in stability (Eq. 4.5). These dimeric proteins have variable ψ_f -values which vary from zero to one according to ρ and the added stability. Folding energy landscapes (*right*) are modeled using ρ and ψ_f -values with two pathways, the N-terminal helix nucleation route containing the biHis site and the C-terminal helix nucleation route. The unfolded ensemble occupies the outer ring. For GCN4^{d,A24G}_{biH3,7} (**a**), the N- and C-terminal barriers are nearly equivalent in the absence of metal, and 50% of the ensemble nucleates at either end of the protein. But at high concentrations of M^{2+} , the N-terminal, metal stabilized route dominates. For GCN4^d_{biH3,7} (**b**), nucleation principally occurs at the C-terminus in the absence of M^{2+} . High concentrations of M^{2+} stabilize the N-terminal barrier to a level comparable to the C-terminal barrier, resulting in nucleation occurring equally at both termini. Graphics were rendered as in Fig. 4.4.

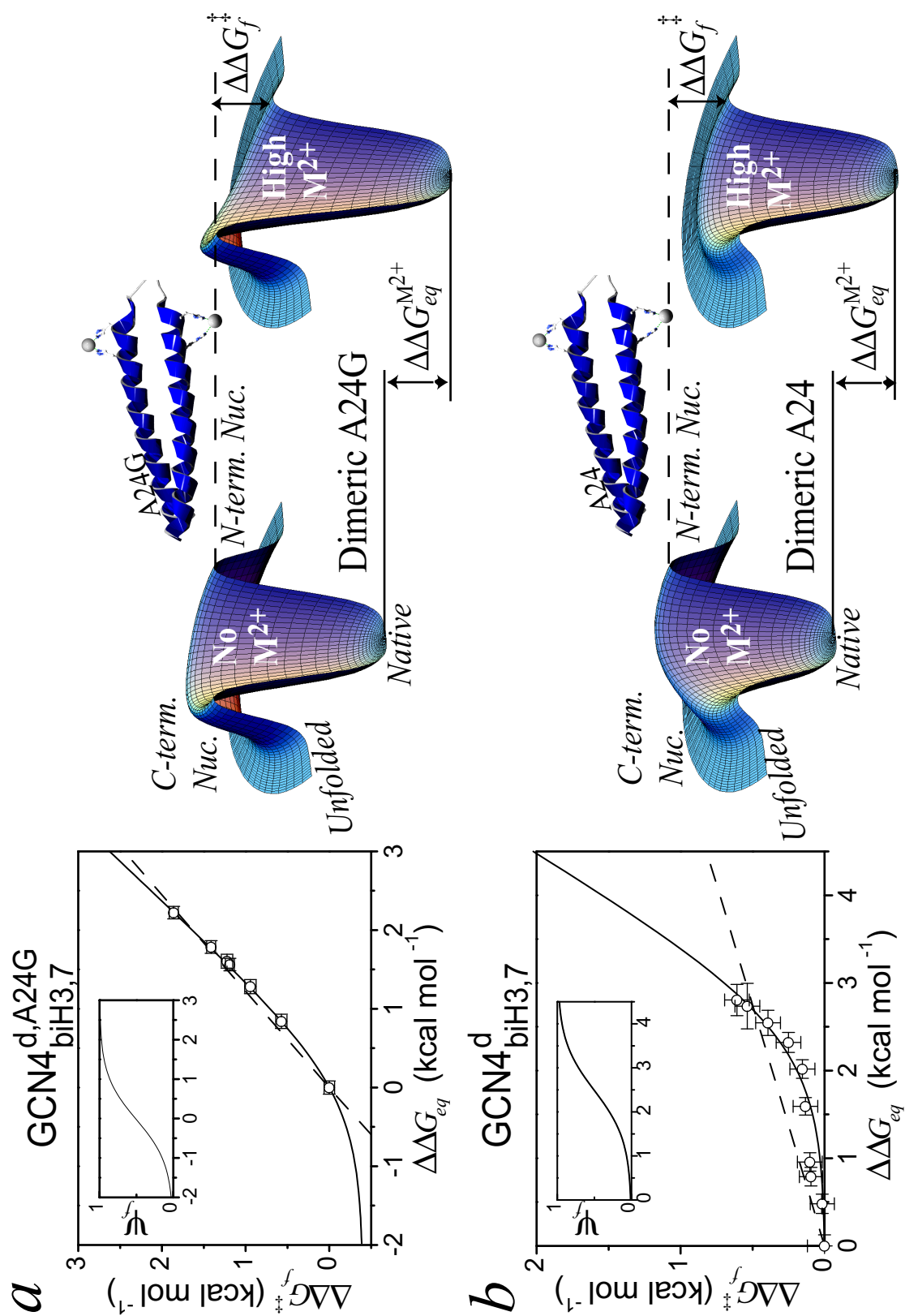


Table 4.3 Heterogeneous and homogeneous folding model fit parameters^a

Protein ^b	Homogeneous Pathway		Heterogeneous Pathways		
	Fraction Formed ϕ_f	Fitness of Model χ^2	Relative Flux ρ	Fraction of Site Formed f	Fitness of Model χ^2
GCN4 ^{Nx,A24G} _{biH3,7}	1.01±0.02	0.15	0.18±0.22	1.05±0.05	0.001 ^c
GCN4 ^{Cx} _{biH3,7}	-0.06±0.02	0.23	N.D. ^d	N.D.	N.D.
GCN4 ^{d,A24G} _{biH3,7}	0.82±0.02	2.31	1.2±0.1	1.00±0.02	0.05
GCN4 ^d _{biH3,7}	0.18±0.02	18.1	83.2±6.2 (63.8±53) ^e	1.00 (0.89±0.2) ^e	0.31 (0.19) ^e

^aCurve fit parameters for Figs. 4.4 and 4.6. Linear fits with fixed zero intercept (i.e.

$\Delta\Delta G_f^\ddagger = \phi_f \times \Delta\Delta G_{eq}$ from Sec. A3.3) are applied for homogeneous models. Eq. 4.4 is applied for heterogeneous models. Models are fit by weighting the inverse of the square of the error bars, yielding χ^2 to compare the appropriateness of each model.

^bGCN4 topologies are depicted in Fig. 4.1, and protein sequences are given in App. 2.

^cWhile χ^2 is smaller for the heterogeneous fit, the lack of curvature in the plot causes ρ to have significant error, and the flux through this pathway is >10-fold. Comparison to the C-terminal linked data indicates that the end nearest the cross-link is favored >1000-fold (Fig. 4.4).

^dNot Determined. Because $\Delta\Delta G_f^\ddagger$ is not affected by metal-induced stabilization up to 3 kcal mol⁻¹, ρ is very large (>1000), but it cannot be determined precisely without significant curvature in the data.

^eFor dimeric GCN4, higher stability data is required to accurately fit ρ and f simultaneously; however, the f -value for dimeric A24G and N-terminal cross-linked A24G versions fit well to $f=1$. The fit with a ρ -value of ~80 when f is fixed to unity is preferred, but the parameters in parentheses are provided, where both values are fit simultaneously.

The second dimeric protein examined is the A24G version, because most nucleation events occur near the N-terminus for this system. At low concentrations of metal, the ψ_f -value is one-half indicating that about one-half of the nucleation events occur in the biHis region (Fig. 4.6a). This uniform distribution occurs despite the fact that the C-terminus has most of the intrinsic helicity (Moran *et al.*, 1999). As the metal concentration is increased, the N-terminal region is further stabilized and nearly all molecules fold with helix nucleated at the N-terminus as evidenced by the ψ_f -value that increases to unity (Fig. 4.6a).

A comparison of dimeric A24 and A24G versions indicates that the change in the degree of pathway heterogeneity recapitulates the difference in their equilibrium stability. The A24G mutation is responsible for a shift in amount of flux going through regions other than the N-terminus, since the ρ -value shifts from $\sim 80:1$ to $\sim 1:1$ (Table 4.3). The ratio of the heterogeneity in these two versions reflects the loss in stability for this mutation.

$$\Delta\Delta G_{eq} = RT \ln (\rho_{Ala}/\rho_{Gly}) \quad (4.6)$$

Thus, the A24G mutation is responsible for the ~ 80 fold change in ρ -value or 2.5 kcal mol⁻¹. This shift is consistent with the decrease in stability for the mutation in either the biHis or wild-type (Moran *et al.*, 1999) backgrounds (1.7 ± 0.1 or 2.4 ± 0.1 kcal mol⁻¹, respectively), implying that the region containing the 24th residue is the major alternative nucleation site (Fig. 4.6).

4.3.5 Properties of the coiled coil transition state

Numerous characterizations of the dimeric and cross-linked forms of the coiled coil indicate that about one-third to one-half of the helix nucleates in the TS. Our initial estimate of one-third helical content resulted from the low ϕ -value of 0.16 for the A14G comparison near the center of the coiled coil (Sosnick *et al.*, 1996a). An H/D backbone amide isotope effect study pioneered by our group (Krantz *et al.*, 2000) quantified the extent of hydrogen bonding in the TS with a $\phi_f^{\text{D-H}}$ -value of 0.6 (Ch. 3). Potentially, the apparent overestimate of the helical content is due to non-native H-bond formation, possibly 3_{10} -helix, or backbone desolvation in the TS, as suggested by our trifluoroethanol studies (Kentsis & Sosnick, 1998) and other recent results (Ibarra-Molero *et al.*, 2001).

In summary, these data indicate that partial desolvation of the polypeptide is a critical element of the rate-limiting step. Desolvation of the backbone is energetically costly, but helix formation offsets this energetic penalty. In order to condense hydrophobic side chains of two amphipathic helices, $i, i+4$ amides and carbonyls form H-bonds, because the clustering of the hydrophobic moieties expels competing water molecules from the amide/carbonyl interface and would leave H-bonding groups unsatisfied in the nascent hydrophobic core. As the initial buried backbone H-bonds form, the solvent exposed H-bonds on the opposite face of the helix form. In this view, pairs of amphipathic helices provide an efficient geometry to cooperatively bury side chain surface while satisfying H-bonding requirements (Krantz *et al.*, 2000; Krantz *et al.*, 2002).

4.3.6 Metal binding and ψ -values

We conclude that metal binding necessitates authentic helix formation. We recover native strength metal binding in the kinetics experiments (Fig. 4.2) and found that the fraction of interaction for the metal binding site, the f -value, was unity. This interpretation is consistent with our D/H isotope effect results (Krantz *et al.*, 2000; Krantz *et al.*, 2002), and other mutational studies (Moran *et al.*, 1999; Zitzewitz *et al.*, 2000). Directly connecting ψ_f -values and helix formation implies that fractional values equate to the degree of pathway heterogeneity (i.e. the fraction of TS's with helix formed at the biHis site). Hence, this method overcomes a number of deficiencies of standard mutational studies. (i) Partial ϕ_f -values either may represent pathway heterogeneity or partial structure formation. (ii) Additional structural perturbations are introduced, especially in core mutations. (iii) Side-chain interactions in the TS often are unknown. (iv) Helix formation is inferential and often complicated by tertiary interactions, or in the case of Ala/Gly comparisons, backbone configurational entropy is probed.

4.3.7 Conclusions

Engineered biHis sites can address questions regarding pathway heterogeneity and helix formation in TS's. By spanning a larger energy range than possible with standard mutations in a manner that does not perturb the structure, we identified minor pathways representing 1% of the flux. Furthermore, the method explores the free energy reaction surface, because a continuous change in metal ion concentration continuously varies the stability of the specific region. The translational symmetry of the dimeric

version of the coiled coil results in multiple helix nucleation sites. Intrinsic helicity of the sequence, however, is not the sole determinant of the site of nucleation. In both the amino- and carboxy-tethered versions, translational symmetry is broken, and folding occurs along a single, dominant pathway that is exclusively determined by tether location. Admittedly, the introduction of a metal binding site in the core of a protein may prove difficult. However, surface metal sites can generally be introduced throughout in many other proteins, thereby permitting investigations of a wide variety of folding issues including the importance of local versus non-local interactions.

4.4 Ubiquitin

To better investigate the degree to which metal ion binding site design could be implemented in the elucidation of folding pathway diversity, I designed metal binding sites in mammalian ubiquitin (Ub). However, because Ub is composed of both α -helical as well as β -sheet secondary structures, metal sites were designed for both types of secondary structures (Fig. 4.7).

4.4.1 *The α -helical sites*

Two separate α -helical divalent metal ion binding sites were designed in ubiquitin using solvent exposed positions in an $i,i+4$ helical geometry. Ub^{H1} contains substitutions E24H and A28H. Site H1 is stabilized by the addition of Co^{2+} by about 2 kcal mol⁻¹. Both equilibrium denaturation melts and kinetic measures confirm that the added stability due to the addition of metal ions is reversible and follows two-state kinetics. In the absence of metal ions, the folding pathways, which include site H1, are formed 38% of the

Figure 4.7 Three-dimensional rendering of designed metal binding sites in ubiquitin. The ribbon diagram shows the location of the designed divalent metal ion binding sites (*yellow balls*) engineered into a fluorescent tryptophan containing version of mammalian ubiquitin (Ub F45W). Sites H1 and H2 are standard $i,i+4$ helical sites identical to the types tested in GCN4. The β -sheet sites, however, do not naturally occur. Sites B1 and B3 bridge the parallel N- and C-terminal strands, whereas site B2 bridges the first anti-parallel β -hairpin. Site T1 is engineered across an Ω -loop structure. In all sites excepting B1, the naturally occurring histidine (H68) was mutated to asparagine. See App. 2 (Table A2.2) for a detailed description of the residues which are mutated for each site.

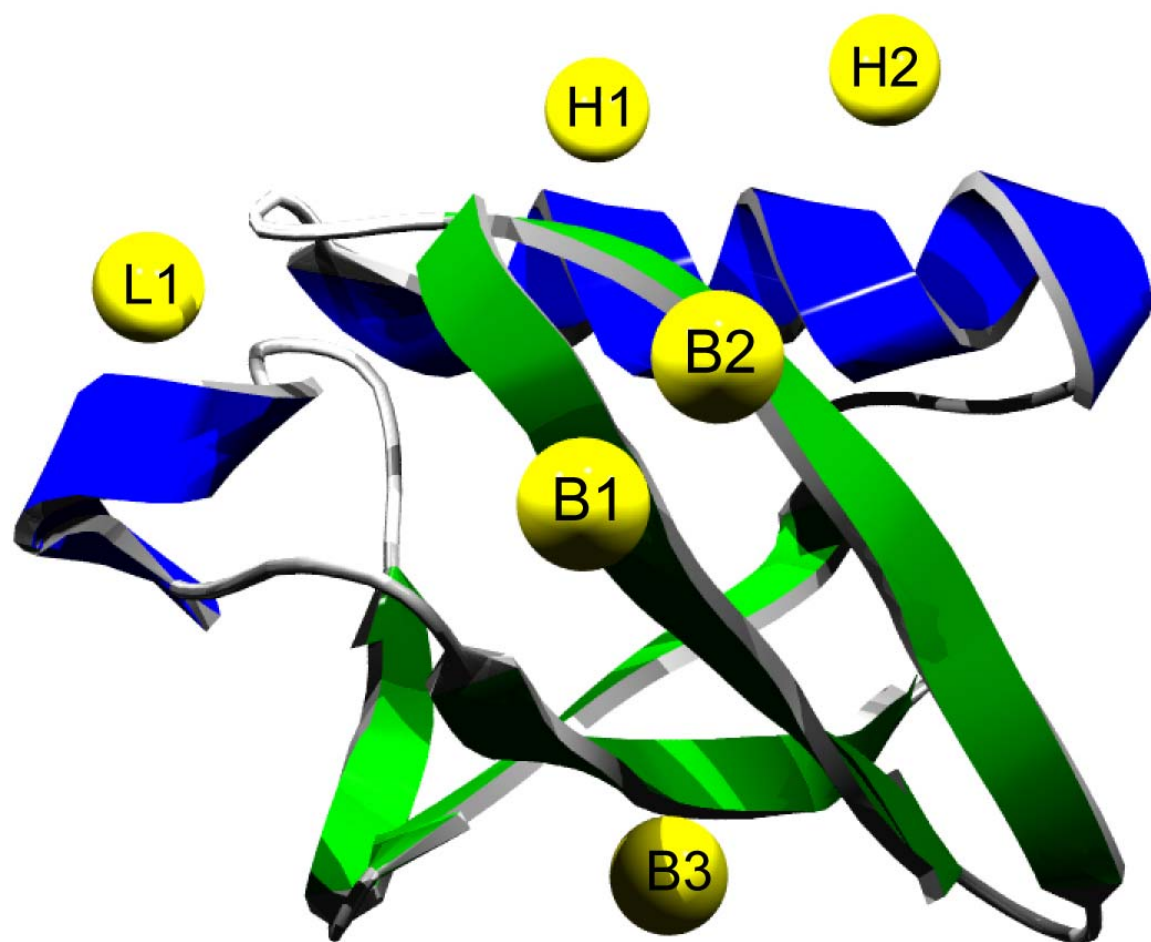
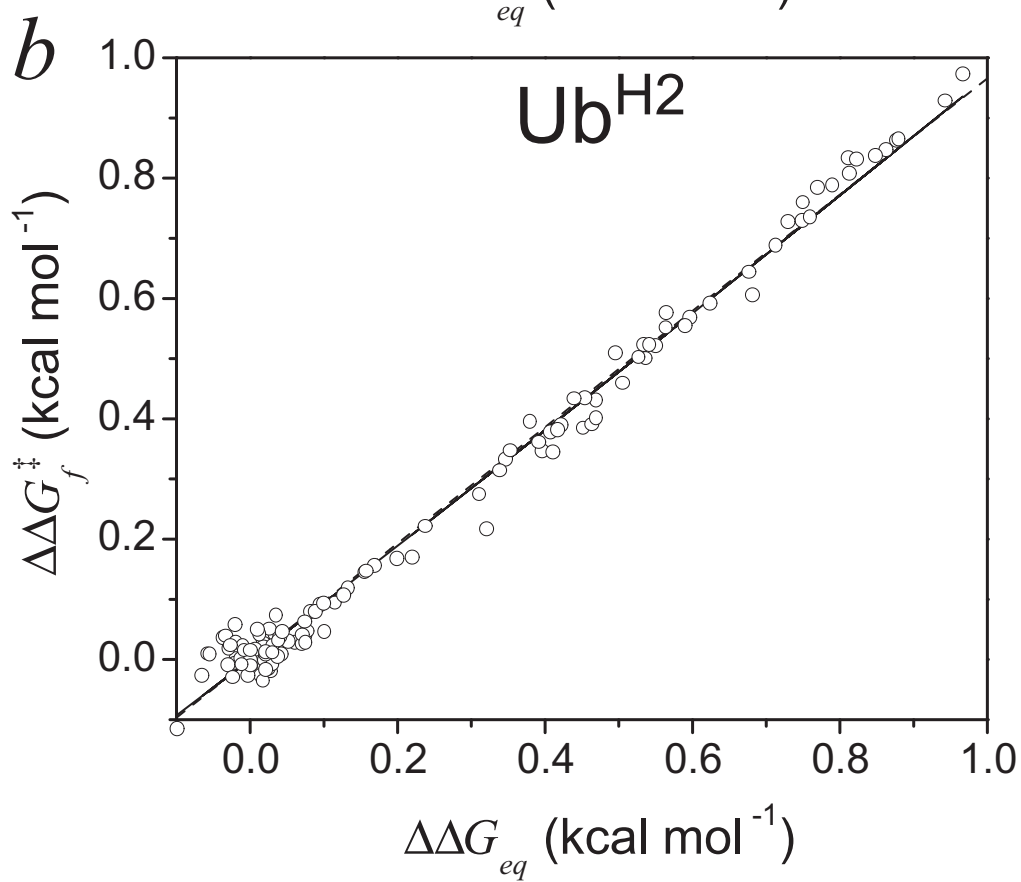
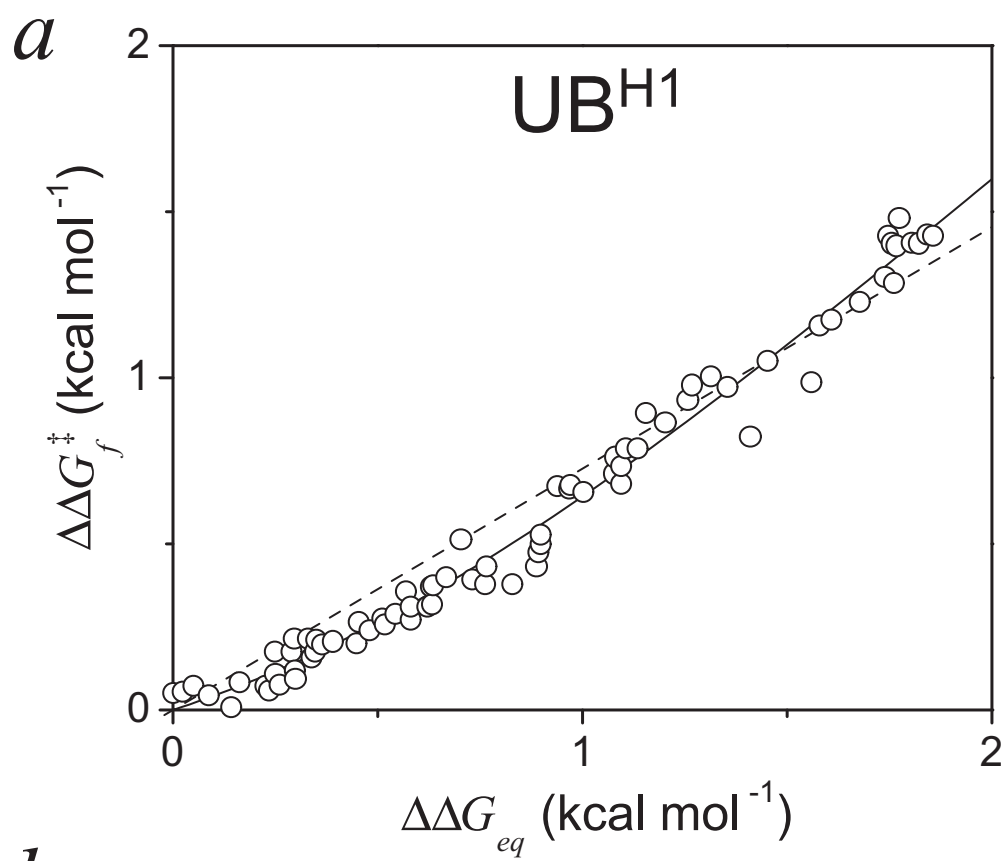


Figure 4.8 ψ -analysis of α -helical sites in ubiquitin. Instead of measuring complete chevrons for each metal ion concentration, as was done for GCN4 (see Sec. 4.3; Figs. 4.3 and 4.5), the metal ion binding work for ubiquitin titrated Co^{2+} at constant denaturant conditions. Thus, the GdmCl concentrations used correspond to either strongly refolding or unfolding conditions. $\Delta\Delta G_f^\ddagger$ is estimated as the difference in folding activation energy of protein in the absence and presence of a particular $[\text{Co}^{2+}]$. $\Delta\Delta G_{eq}$ is calculated from the analogous unfolding activation energy, $\Delta\Delta G_u^\ddagger$, according to $\Delta\Delta G_{eq} = \Delta\Delta G_u^\ddagger - \Delta\Delta G_f^\ddagger$. **a**, Ub^{H1} was refolded and unfolded in 1.0 M and 6.3 M GdmCl, respectively. **b**, Ub^{H2} was refolded and unfolded in 1.5 M and 6.14 M GdmCl, respectively. ψ -analysis fits (—) are given according to Eq. 4.4, and simple ϕ -analysis fits (--) are given as well. Fit parameters for ψ -analysis fits are given in Table 4.5.



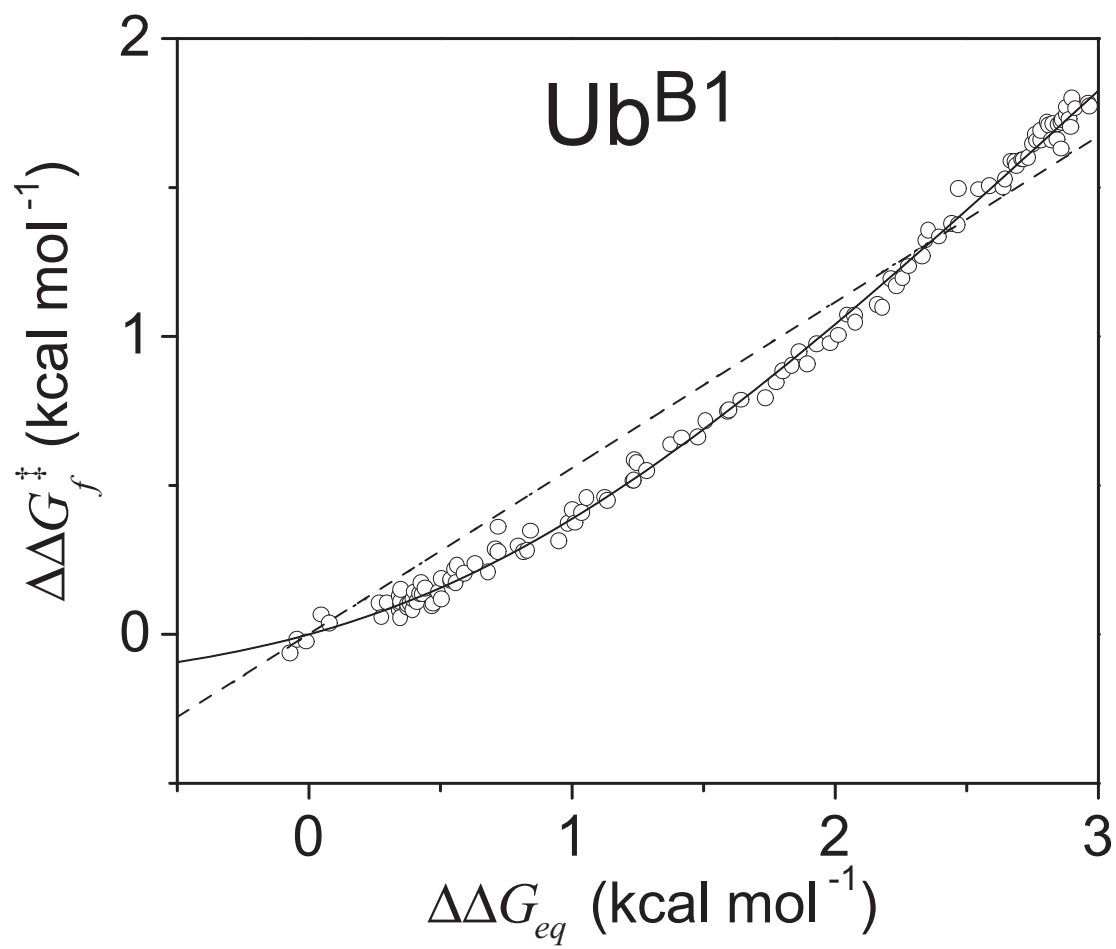
time (a ρ -value of 1.64 ± 0.02 ; Fig. 4.8; Table 4.5). However, when the concentration of Co^{2+} is increased, the stabilization of this pathway increases the flux through this pathway until it is fully populated. Finally, the fraction to which the metal site itself is formed (f -value) in the rate-limiting step is near unity (1.06 ± 0.03) as is observed in the coiled coil, GCN4 (Table 4.5).

The second protein, Ub^{H2} , was created by the double substitution of A28H and D32H. This site afforded less added equilibrium stability than Ub^{H1} although equilibrium and kinetic measures confirmed that the change in added stability due to metal ion binding is a reversible process for a simple two-state folding model. Also, kinetic studies revealed that the ρ -value was low (0.07 ± 0.01), demonstrating that this portion of the helix is crucial to the rate-limiting ensemble (Fig. 4.8). In 93% of the rate-limiting structures, helix occurs at site H2 (Table 4.5). This fact may seem to conflict with the fact the site H1 is less well formed; however, site H2 was initially destabilized by the introduction of the histidine residue (Table 4.4). Once correcting for this destabilization, the relative flux through either end of the helix is more consistent, albeit slightly more so at site H2. Ubiquitin's helix is present in nearly all rate-limiting species, which is consistent with the high ϕ -values found there (mean ϕ_f of 1.27; Table 4.4). Nevertheless, a small percentage ($\sim 10\%$) of transition state species do not require the helix.

4.4.2 The β -sheet sites

As proteins are composed of both helical and sheet secondary structures, designed metal ion binding sites were incorporated into ubiquitin's four stranded β -sheet network

Figure 4.9 ψ -analysis of a β -sheet site in ubiquitin. $\Delta\Delta G_f^\ddagger$ and $\Delta\Delta G_{eq}$ are estimated as described in Fig. 4.8. Ub^{B1} was refolded and unfolded in constant 1.72 M and 5.76 M GdmCl, respectively. Zn²⁺, however, was selected as the stabilizing metal ion. ψ -analysis fits (—) are given according to Eq. 4.4, and simple ϕ -analysis fits (--) are given as well. Fit parameters for ψ -analysis fits are given in Table 4.5.



in both parallel and anti-parallel topologies (Table A2.2; Fig. 4.7). Since these sites are more unconventional than the $i,i+4$ helical sites which are found naturally in zinc finger proteins, it was imperative that these sites were tested for perturbation of the native structure with and without the presence of divalent metal ions. Of the three sheet sites tested, site B1 and B3 contained the same circular dichroism spectra under native conditions with and without metal ions. However, Ub^{B2} possessed less secondary structure than wild-type in the absence of metal ions, where the addition of metal ion increased the amount of ellipticity in the native structure. Thus, in contrast to the more natural and robust helical sites, only two-thirds of the β -sheet sites were non-invasive to the native secondary structure.

Metal binding kinetic ψ -analysis has thus far been applied to Ub^{B1}. Here this parallel β -sheet site shows considerable heterogeneity with a ρ -value of 2.4 ± 0.1 (Fig. 4.9; Table 4.5). Therefore, the parallel β -sheet is formed in 29% of the rate-limiting ensemble. The addition of Zn^{2+} increases the probability of nucleation at the parallel sheet site demonstrating the plasticity of the rate-limiting ensemble.

4.4.3 Secondary structure metal ion preferences

Having now studied both helical and sheet metal ion binding sites, some interesting trends have developed regarding their respective metal ion preferences. With the metric of the maximal stabilization at saturating concentrations of metal ion in mind, it was determined that Co^{2+} (2.8 to 3.2 kcal mol⁻¹) provides about twice as much stability as Zn^{2+} (1.2 to 1.5 kcal mol⁻¹) in *helical* sites (Fig. 4.2; Table 4.1). Whereas the

opposite is true for the parallel β -sheet site, which prefers Zn^{2+} (2.9 kcal mol⁻¹) to Co^{2+} (0.8 kcal mol⁻¹). The most logical conclusion to this apparent switch in metal ion preference is related to the geometric differences inherent to helix and sheet. Also Zn^{2+} and Co^{2+} prefer differing coordination geometry. While Zn^{2+} is generally found in tetrahedral coordination geometry, Co^{2+} often maintains octahedral geometry. In terms of bond angles, the tetrahedral geometry requires liganding histidines to bind at 109.5° while Co^{2+} requires 90° between liganding histidines. Backbone ϕ, ψ -angles restrict the χ -angles accessible to the histidines' imidazole side chain moiety; therefore, future metal ion site design may better predict the engineered sites and metal ion best suited for each site.

4.4.4 Conclusions for ubiquitin

Experiments described here for ubiquitin have tested whether metal ion binding sites may be engineered into a more complex α/β -fold. While the helical sites are robust and have been successful in the coiled coil and Ub, one of the anti-parallel β -sheet sites engineered into ubiquitin significantly perturbed the secondary structure prior to the addition of metal ion. However, a parallel site engineered into ubiquitin has secondary structure consistent with the native protein and allowed for kinetic characterization. The primary goal of metal ion site ψ -analysis is to assess folding pathway heterogeneity. To this end, the helical sites are pre-requisite and are uniformly required in the rate-limiting ensembles, whereas the parallel β -sheet site is only required in 29% of the transition state ensemble. This method, therefore, is the first to identify pathway heterogeneity in parallel β -sheet as well as quantify the heterogeneity.

Table 4.4 Chevron analysis of amino acid substitutions for Ub metal site variants

Protein ^a	$-\Delta G_{kinetics}$ (kcal mol ⁻¹) ^b	$RT \ln k_f^{H_2O}$ (kcal mol ⁻¹) ^c	$m_{kinetics}^o$ (kcal mol ⁻¹ M ⁻¹)	m_f (kcal mol ⁻¹ M ⁻¹)	ϕ_f
Ub	7.50±0.15	2.69±0.06	2.27±0.04	1.45±0.03	N.A.
UbA28G	7.04±0.13	2.18±0.06	2.33±0.04	1.52±0.04	1.1 ^d
Ub ^{pWT}	7.20±0.12	2.82±0.07	2.44±0.04	1.64±0.04	-0.4 ^d
Ub ^{H1}	6.60±0.11	1.82±0.07	2.37±0.04	1.58±0.04	1.67 ^e
Ub ^{H2}	6.70±0.11	2.30±0.07	2.41±0.04	1.66±0.04	1.04 ^e
Ub ^{B1}	6.39±0.08	2.78±0.04	2.11±0.03	1.35±0.03	-0.08 ^e
Ub ^{B2}	6.36±0.07	2.54±0.05	2.25±0.03	1.48±0.03	0.13 ^e

^aSee App. A2 for details of the individual amino acid substitutions for variant (Table A2.2).

^bFolding equilibrium free energy estimated from folding kinetics in the absence of denaturant; the refolding buffer is 50 mM HEPES, pH 7.5, 1 mM EDTA, 10 °C.

^cFolding activation energy in the absence of denaturant.

^dFolding ϕ -values are referenced to the Ub background.

^eFolding ϕ -values are referenced to the Ub^{pWT} background.

Table 4.5 Kinetic folding pathway parameters from ψ -analysis^d for ubiquitin

Protein	f	ρ_{local}^a	Local flux ^{a,c}
Ub ^{H1}	1.00 ^b (1.06±0.03)	1.64±0.02 ^b	38%
Ub ^{H2}	1.00 ^b (0.965±0.007)	0.07±0.01 ^b	93%
Ub ^{B1}	0.834±0.008	2.4±0.1	29%

^a“Local” refers to parameters estimated within each individual biHis variant’s background. A “global” analysis would be required to rigorously compare the individual variants using the same “wild type” Ub background.

^bFrom previous work on GCN4 helical sites have f -values of unity (see Sec. 4.3); therefore, ρ -values are determined when f is fixed to unity. Values in parentheses are the best fit f -values.

^cFlux calculated as $1/(1 + \rho)$ which is the percentage of the ensemble that is native-like in the biHis probed region at the rate-limiting step.

^dSee Sec. 4.3.6 (Eqs. 4.4 and 4.5) for a full description of ψ -analysis.

The pseudo-rotational fold symmetry in ubiquitin may account for some of the heterogeneity. As in GCN4 coiled coil, symmetry in the fold creates the inherent possibility of multiple folding routes, since topology dominates the rate-limiting step of folding. Only local stabilities may bias symmetrical nucleation sites in ubiquitin. However, these local biases are too weak to prevent substantial heterogeneity in the parallel β -strand.

4.5 Other studies of pathway heterogeneity

For two-state proteins other than GCN4, limited evidence supports the possibility of pathway heterogeneity. For example, the topologically identical, small IgG binding domains of protein G and protein L nucleate at either of two different, but symmetrically related, turns (Grantcharova *et al.*, 2001). Mutational studies indicate that nucleation can occur at either turn and reflects their relative stability. Although mutational ϕ -analysis can identify multiple pathways, both the Serrano (Martinez *et al.*, 1998; Martinez & Serrano, 1999) and Baker (Grantcharova *et al.*, 2000) groups found no evidence for a shift in pathway upon the destabilization of elements formed in the TS of SH3 domains. Similarly, loop insertion studies did not detect pathway heterogeneity (Grantcharova *et al.*, 2000; Viguera & Serrano, 1997). However, homologs of SH3 fold through different TS's (Grantcharova *et al.*, 2001; Guerois & Serrano, 2000).

When Fersht *et al.* encountered the ambiguity of assessing pathway heterogeneity in chymotrypsin inhibitor 2 (CI2) folding studies (Fersht *et al.*, 1994), they concluded that a homogenous pathway could explain the data. Baker *et al.*, however, reanalyzed the solvent exposed portion of the CI2 data, concluding that the ambiguity persists regarding

whether partial ϕ -values are due to pathway heterogeneity (Kim *et al.*, 1998). Their analysis conceded that mutagenesis methods are insufficient to generate the ~ 4 kcal mol⁻¹ free energy change required to establish the appropriate folding model for ProtL.

Recent work has investigated the role of topology either through cross-linking or circular permutation. Topological changes sometimes (Grantcharova & Baker, 2001; Hennecke *et al.*, 1999; Viguera *et al.*, 1996) but not always (Grantcharova *et al.*, 2000; Otzen & Fersht, 1998) result in different TS's. The existence of distinct TS's reveals that there is structural diversity at the rate-limiting step. However, such methods neither imply that heterogeneity exists for the *same* version of the protein nor do they quantify the diversity of TS's.

Proteins that fold in a two-state manner are well-suited for developing single molecule methods. However, single molecule methods will not detect pathway heterogeneity. Unfolded ensembles typically equilibrate faster than the folding rate. Only a single relaxation process is observable, with a rate being the sum of the rates for all pathways. Thus, individual routes are not identifiable, unlike measurements on systems that have kinetically distinct starting configurations and exhibit multi-exponential relaxation processes (Zhuang *et al.*, 2000).

4.6 Methods

4.6.1 Proteins

3,7-biHis versions of GCN4-p1' (Ac-RMHQLEHKVEELLSKNWNLENEVAR LKKLVGER-NH₂) were synthesized with Y17W for fluorescence. Cross-linked versions (Moran *et al.*, 1999) contain either an additional CGG or GGC linker at the N- or C-

terminus, respectively. For further detail on the protein sequences see App. 2. Dimeric versions lack the terminal cysteine residue, preventing unwanted oxidation at pH 7.5. Peptide concentrations are determined using $\epsilon_{280}=5,700 \text{ M}^{-1} \text{ cm}^{-1}$. Mass spectrometry confirmed all protein samples.

Ubiquitin mutants were created using the Stratagene Quik Change Kit. Exact amino acid substitutions are described in App. 2 (see Table A2.2). Proteins were expressed and purified according to App. 2.

4.6.2 Equilibrium

Equilibrium stability was determined from GdmCl denaturation using CD₂₂₂ at 2 nm resolution. Peptide concentrations ranged from 2 to 42 μM , in Buffer A (20 mM HEPES, 150 mM NaCl at pH 7.5), 10 °C. EDTA was added at 0.5 to 1 mM final concentration for ubiquitin variant chevrons as a scavenger of metal ion contamination.

4.6.3 Stopped-flow kinetics

Experiments used a Biologic stopped-flow apparatus (Krantz & Sosnick, 2000) at 10 °C in Buffer A. Protein concentrations ranged from 0.1-32 μM . Metal ions stock solutions (1.0 M CoCl_2 or 0.25 M ZnCl_2) were added in equal concentrations to buffers.

5

Conclusions and Future Directions

5.1 General conclusions

Analysis in the second chapter shows that the folding of small globular proteins appears almost universally to be kinetically two-state, in spite of the prevalence of on-pathway intermediates. The Initial Barrier Hypothesis anticipates this otherwise surprising result (Sosnick *et al.*, 1996b). The hypothesis proposes that proteins enter the folding pathway only when the search through U-state space chances to find a topologically native-like conformation that can serve as a transition state able to nucleate downhill native-like structure formation. Stable native-like intermediates essentially do not form before the initial large scale search step succeeds. Stable native-like intermediates do not accumulate afterwards, because later intrinsic barriers represent smaller scale conformational searches that are not rate-limiting. Exceptions to this generalization result from adventitious or imposed misfolding events.

In the third chapter, I describe the development of procedures that determine when hydrogen bonds form under fast, two-state folding conditions using equilibrium and kinetic deuterium/hydrogen isotope effects in backbone amides. A library of model proteins is analyzed by this method. The equilibrium isotope effect is quantified by the “ C_m experiment,” upon which the change in equilibrium stability of a protein is monitored usually by far UV circular dichroism in a continuous fashion during hydrogen

exchange. Stopped-flow kinetics experiments quantify not only the kinetic isotope effect of deuterating backbone amides but also the equilibrium isotope effect. The change in equilibrium free energy accompanied by backbone deuteration correlates to secondary structure composition when hydrogen bonds are divided α -helical and β -sheet classifications. The number of helical hydrogen bonds correlates with an average equilibrium isotope effect of $8.6 \pm 0.9 \text{ cal mol}^{-1} \text{ H-bond}^{-1}$ where protium is more stabilizing than deuterium. However, β -sheet and long range H-bonds have little, if any, isotope effect. I then calculate the contribution of the hydrogen to protein stability using a scale factor, finding that the contribution of hydrogen bonds varies due to the inherent differences of helical and sheet site isotope effects. The origin of these differences, I speculate, derives from the fact that sheet hydrogen bonds are more diverse in ϕ, ψ orientations than helical ones.

The global kinetic isotope effects supports the hypothesis that hydrophobic association cannot be divorced from helix formation in the rate-limiting step. Extensive backbone desolvation is a critical component of the TS, with helix formation providing an efficient means to satisfy the H-bond requirements of hydrophobically collapsed environment. Thus folding models describing an incremental build-up of structure where hydrophobic burial and H-bond formation occur commensurately are more consistent with these data than are models which posit the extensive pre-formation of either quantity in isolation.

In the fourth chapter, I showed that engineered biHis sites may address questions regarding pathway heterogeneity and α -helix and β -sheet formation in transition states. I

am able to span a larger energy range than possible with standard mutations in a manner that does not perturb the structure as significantly as traditional methods, wherein multiple amino acids are substituted at a particular site. Minor pathways representing 1% of the flux in GCN4 coiled coil are identified. Furthermore, the method explores the free energy reaction surface, because a continuous change in metal ion concentration continuously varies the stability of the specific region.

Translational symmetry in the dimeric version of the coiled coil results in multiple helix nucleation sites. Intrinsic helicity of the sequence, however, is not the sole determinant of the site of nucleation. In both the amino- and carboxy-tethered versions, translational symmetry is broken, and folding occurs along a single, dominant pathway that is exclusively determined by tether location.

Experiments on ubiquitin tested whether metal ion binding sites may be engineered into a more complex α/β -fold. The central goal of metal ion site ψ -analysis is to assess folding pathway heterogeneity. I found that the helical sites are pre-requisite and uniformly required in the rate-limiting ensembles (up to 93%), whereas the parallel β -sheet is only required in 29% of the transition state ensemble. This novel method, therefore, is the first to identify pathway heterogeneity in parallel β -sheet and quantify it.

The pseudo-rotational fold symmetry in ubiquitin may account for some of the heterogeneity. As in GCN4 coiled coil, symmetry in the fold creates the inherent possibility of multiple folding routes, since topology dominates the rate-limiting step of folding. Only local stability may bias symmetrical nucleation sites in ubiquitin. However, these local biases are too weak to prevent substantial heterogeneity in the parallel β -strand.

5.2 Future directions

Future protein folding studies should better interface with theory. First, it is widely established that proteins fold in an apparent two-state fashion. Second, both a surface area burial and H-bond reaction coordinate have been established for helical proteins; these coordinates can be rigorously tested in folding computer simulations. The interplay of simulation and these experimental data will allow for continual refinement of the energetic parameters used in simulations. Finally, the advent of metal binding ψ -analysis enables a powerful constraint to be placed on the diversity of the folding pathways. Metal binding ψ -analysis will allow folding simulation parameters to be assessed rigorously. For example, folding simulations that favor rate-limiting structures that are less frequent as determined experimentally can now be thoroughly critiqued.

Since kinetic determination of rate-limiting β -sheet structure has remained elusive to backbone amide isotope effect studies. A strong future remains for metal binding ψ -analysis; however, β -sheet sites prove to be more difficult to design than their α -helical counterparts. Thus, multiple rounds of mutagenesis at particular β -sheet sites in combination with computer modeling may allow for this potentially important area of protein folding pathway heterogeneity to be better understood.

Appendix 1

Protein Expression Vectors and Mutants

Table A1.1 Expression vectors from total DNA synthesis

Protein ^a	Expression Plasmid ^b	Synthetic DNA Sequence ^c
chymotrypsin inhibitor 2 (CI2)	pRS.CI2	CATATGAAGACAGAGTGGCCAGAGCTCGTGGGGAAATCGGT GGAGGAGGCCAAGAAGGTGATTCTGCAGGACAAGCCAGAGG CGGAAATCATAGTTCTGCCGGTGGGTACCATTGTGACCATG GAATATCGGATCGATCGCGTCCGCCTCTTTGTGCGACGATCT AGACAACATTGCCCAGGTCCCGCGGGTCGGCTAAGCTT
common type acylphosphatase (CTACP)	pRS.CTACP	CATATGGCAGAGGGAAACACCCTGATATCAGTGGATTATGA AATTTTTGGGAAGGTGCAAGGGGTGTTTTCCGTAAGCATA CTCAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTC CAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGG TCCAATCTCCAAGGTGCGTCACATGCAGGAATGGCTTGAAA CAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACCTTC AACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTT CCAAATTGTAAAATAAGCTT
colicin immunity protein 7 (IM7)	pRS.IM7	CATATGGAAGTAAAAATAGTATTAGTGATTACACAGAGGC TGAGTTTGTTCACCTTCTTAAGGAAATTGAAAAAGAGAATG TAGCTGCAACTGATGATGTGTTAGATGTGTTACTCGAACAC TTTGTAATAAATTACTGAGCATCCAGATGGAACGGATCTGAT TTATTATCCTAGTGATAATAGAGACGATAGCCCCGAAGGGA TTGTCAAGGAAATTAAAGAATGGCGAGCTGCTAACGGTAAG CCAGGATTTAAACAGGGCTAAGCTT
monomeric lambda repressor protein CI, residues 6-85, G46A, G48A (λ^{AA})	pRS.LambdaAA	CATATGTCGTTAACACAAGAGCAGCTCGAGGACGCACGTGCG ACTTAAAGCAATTTGGGAAAAAAGAAAAATGAGCTCGGCT TATCCCAGGAATCTGTGCGAGACAAGATGGGGATGGGGCAG TCTGCAGTTGCTGCTTTATTTAATGGCATCAACGCGTTAAA TGCTTATAACGCCGCATTGCTAGCAAAGATCTTAAAGTTA GCGTTGAAGAATTTCAGCCCATCGATCGCCCGGGAAATCCGA TAAGCTT
scr-SH3 domain from avian sarcoma virus (SH3)	pRS.SH3	CATATGGGATCCAACATGGGCGGCGTCACCACTTTTCGTGGC TCTCTACGACTACGAGTCCCGGACTGAAACAGATCTGTCCT TCAAGAAAGGAGAACGCCTGCAGATTGTTAACACACGGAA GGTGACTGGTGGCTAGCTCATTCCCTCACTACAGGACGGAC GGGCTACATCCCCAGTAACTATGTGGCGCCCTCAGACTAAG CTT

^aMore information on these proteins is provided in Ch. 3 (Table 3.1, Fig. 3.1, Fig. 3.2) and App. A2

^bProtein expression plasmid constructs were created by ligating into pRSUB at unique 5' Nde I and 3' Hind III restriction sites. pRSUB was a kind gift of K. D. Wilkinson (Emory University).

^cTotal DNA synthesis generated these sequences with optimal codon usage and restriction enzyme placement (Operon Technologies).

Table A1.2 Expression vectors and mutants

Protein ^a	Mutation ^b	Expression Plasmid
CI2		pRS.CI2
CTACP		pRS.CTACP
IM7		pRS.IM7
λ^{GG}		pRS.LambdaGG
λ^{AA}	G46A, G48A	pRS.LambdaAA
SH3		pRS.SH3
ProtG	6-His tag	pET.H6-ProtG ^c
ProtL	Y47W, 6-His tag	pET.H6-ProtL-Trp ^c
Ub ^{WT}		pRS.Ub ^d
Ub	F45W	pRS.Ub.F45W
Ub ^{A28G}	F45W, A28G	pRS.Ub.F45W.A28G
Ub ^{pWT}	F45W, H68N	pRS.Ub.F45W.H68N
Ub ^{H1}	F45W, H68N, E24H, A28H	pRS.Ub.F45W.H68N.E24H.A28H
Ub ^{H2}	F45W, H68N, A28H, D32H	pRS.Ub.F45W.H68N.A28H.D32H
Ub ^{B1}	F45W, K6H	pRS.Ub.F45W.K6H
Ub ^{B2}	F45W, H68N, Q2H, E16H	pRS.Ub.F45W.H68N
Ub ^{B3}	F45W, H68N, Q2H, E64H	pRS.Ub.F45W.H68N.Q2H.E64H
Ub ^{T1}	F45W, H68N, R54H, D58H	pRS.Ub.F45W.H68N.R54H.D58H

^aAdditional information on these proteins is provided in Chs. 3 and 4, Table A1.1, and App. A2.

^bAll mutations made with Stratagene Quik Change kit, excepting ProtL.

^cSix His tagged Protein G and Protein L expression vectors were a kind gift of D. Baker (U. Washington).

^dThe wild type mammalian ubiquitin expression vector was a kind gift of K. D. Wilkinson (Emory U.) pRS.Ub is derived from the pRSET B expression vector (Invitrogen) and a chemically synthesized clone of ubiquitin.

Appendix 2

Protein Sequences and Purification Methods

A2.1 General protein expression and purification methods

This appendix is not meant to replace existing manuals on protein expression and purification, but rather it is designed as a guide for not only the proteins encountered in this thesis but also for many other small globular proteins used in similar biophysical studies. Essentially, these are general procedures that are sometimes called “the tricks of the trade” or the unpublished secrets that make protein purification more reproducible.

A2.1.1 Expression cell lines and competency protocols

When selecting a particular type of expression cell line, keep in mind past performance. I have found that Novagen BL21(DE3) cells outperform Stratagene’s expression cell line. Results will vary and depend on inclusion body formation or codon usage in your protein’s DNA sequence, especially rare arginines. With all that said, I always use the Novagen cells. Once a stock of cells is purchased, competent cells can be made indefinitely by the CaCl_2 procedure outlined below.

Step 1: Take an 100 μL aliquot of frozen cells from the -80°C and inoculate about 500 ml to 1 L sterile LB broth. Do not add antibiotic, since these cells do not have a plasmid in them. Work as sterile as possible. *Step 2:* Grow the cells on a shaker at 37°C until they reach an $\text{OD}_{600\text{nm}}$ of 0.3 to 0.4 (1 cm path length; reference the

spectrophotometer to air). *Step 3*: Centrifuge in the Sorvall GSA rotor (250 ml centrifuge bottle) at 5,000 RPM for 10 minutes at 4°C. Ice down 100 mM CaCl₂ and 100 mM MgCl₂ solutions at this point. *Step 4*: Gently resuspend the bacteria pellet on ice in 1/4 volume of ice cold MgCl₂, taking 3 to 5 minutes for this procedure. Centrifuge the cell suspension at 4,000 RPM in the Sorvall GSA rotor for 10 minutes. *Step 5*: Resuspend the bacteria pellet on ice in 1/20 volume of ice cold CaCl₂ and then add an additional 9/20 volume of CaCl₂. Keep this suspension on ice for at least 20 minutes. *Step 6*: Centrifuge the cell suspension at 4,000 RPM in a GSA rotor for 10 minutes and resuspend the cell pellet in 1/50 volume of ice cold, sterile 85 mM CaCl₂ in 15% glycerol w/v. Dispense in 100 µL aliquots and freeze cells immediately at -80°C.

A2.1.2 Protein expression and lysis

Step 1: Transform appropriate DNA plasmid into BL21(DE3) *E. coli* cells. These cells must be competent (see A2.1.1). **(a)** Take a competent cell stock aliquot (about 100 µL) out of -80°C freezer. Place on ice to thaw about 5 minutes. **(b)** Take one µL of plasmid DNA and add to the thawed competent cells. Vortex briefly to mix, and let incubate on ice for at least 20 minutes. Longer incubation will not affect results. **(c)** Heat shock the competent cells at exactly 42°C for 1 minute. Place on ice. **(d)** Add heat shocked cells to 1 ml of LB broth in a sterile culture tube, but do not add antibiotic. Shake at 37°C for at least 1 hour. Shaking for 3 hours does not affect the results. **(e)** Plate the cell culture onto LB agar plates (usually 100 µL of culture in adequate, although it depends on the competency of the cells and the initial concentration of plasmid). Make sure the plates have the appropriate antibiotic (usually ampicillin at 100 µg/ml

concentration). Incubate the plate at 37°C for 14 hours (overnight generally). Longer times of incubations will allow for satellite colonies to appear.

Step 2: Make a starter culture for protein expression. Usually to about 250 to 500 ml of LB broth the antibiotic is added. **(a)** Generally, the antibiotic is ampicillin, which is added to the LB broth at 100 µg/ml final concentration. (Use a stock of ampicillin stored in the 4°C, which is 1000× concentrated or 100 mg/ml in sterile distilled water). **(b)** Pick a colony or two from your plate using a sterile inoculation loop and add to the LB broth/ampicillin flask. Incubate in a shaker (250 RPM) at 37°C until the OD_{600nm} is 0.5 to 1.5 OD. (Anywhere in this range is fine, however, extremely dense starter cultures >1.5 OD can cause problems in protein expression. Under these dense growth conditions the plasmid seems to be rejected or turned-off by BL21(DE3)s.) **(c)** Save your starter culture flask by placing in the 4°C overnight. This will slow down the bacterial growth enough, but prevent the problems dense cultures cause as listed above.

Step 3: Make the big batch of bacteria culture for protein expression. (Expect about 100 mg of protein per liter, but that estimate is very dependent on the protein.) **(a)** Equally divide your starter culture among several 1.5 L volumes of sterile LB broth in 2.8 L Fernbach culture flasks. (Good precision when equally dividing the starter will better control the timing when induction may occur.) **(b)** Add antibiotic to 100 µg/ml final concentration. Incubate in a New Brunswick shaker large enough to hold several Fernbach flasks at 250 RPM. **(c)** Grow until an OD_{600nm} of 0.8 to 1.2. Then induce the culture to express protein by adding 0.3 mM IPTG (isopropylthiogalactoside) or ~0.1 gram per 1.5 liter flask. Only measure out exactly the amount of IPTG you need for your flasks. Dissolve that amount in about 10 ml of sterile water and divide it equally among

your 1.5 L flasks. This stage is called induction. Keep the culture shaking at 37°C. **(d)** Induction of protein generally takes 3 to 4 hours, (but this too depends on the protein). After induction, centrifuge your bacteria in 500 ml bottles in the Sorvall SLA-3000 rotor at 5,000 RPM. Do this in batches until all your culture is spun down saving the cell pastes each time. Freeze the cell paste at -80°C.

Step 4: Lysis and sonication of the bacteria. There are many ways to lyse bacteria, but this method is sufficient. **(a)** Make Lysis Buffer: 25 mM TRIS-Cl, 2 mM EDTA, pH 7.6. (Protease inhibitors like benzamidine or PMSF may be included, but add these right before lysing. Make concentrated PMSF solutions in pure ethanol, since it is hydrophobic.) Then add lysozyme at 100 µg/ml concentration (or a spatula tip). Generally the volume of lysis buffer is 1/20 to 1/50 the volume of the bacterial culture. **(b)** Resuspend the frozen cell paste as best you can in the Lysis Buffer using a 10 ml pipette or whatever means necessary. Let this suspension incubate for 20 minutes at room temperature, or until the suspension becomes turbid and viscous due to release of the bacteria's genomic DNA. **(c)** In order to eliminate the extreme turbidity of the suspension, sonicate the suspension to shear the DNA until the turbidity is similar to that of a concentrated protein solution. Then centrifuge at 18,000 RPM in a Sorvall SS-34 rotor (or $\sim 40,000 \times g$) for 20 minutes at 4°C. **(d)** Save the pellet and the supernatant (referred to as Lysis Supernatant). (If the solution is slightly turbid due to residual intact genomic DNA, a quick way to shear the DNA is to pass through a 22 gauge needle.) Generally, the protein is in the supernatant, but freeze pellets until the protein fraction is found. Protein purification from this supernatant, however, will depend on the properties of the protein: its isoelectric point (pI), size, hydrophobicity, etc. (see methods below).

A2.1.3 Acetic acid precipitation

Acetic acid precipitation is a common step I use which largely eliminates many of the bacteria's cellular proteins found in the Lysis Supernatant. Slowly add concentrated acetic acid drop-wise to the Lysis Supernatant until the solution becomes milky white and the pH of the solution is around 4.8. Centrifuge at 18,000 RPM in a Sorvall SS-34 rotor (or $\sim 40,000 \times g$) for 20 minutes at 4°C. Retain the cell pellet and supernatant. Generally small globular proteins tolerate this step quite well and remain in the supernatant, but results vary for larger proteins.

A2.1.4 Ion exchange chromatography

Ion exchange is the workhorse of protein purification. I exclusively use two resins from Pharmacia. Their brand name for cation exchange is "Fast flow SP" which is an abbreviation for "sulfopropyl" (a charged sulfate ion attached to a sephadex matrix with a three carbon spacer). The anion exchange product is called "Fast Flow Q" which is quaternary amine, creating a single positive charge on a nitrogen and is attached similarly to the SP resin. Before choosing a resin, always calculate the target protein's isoelectric point (pI). Follow this general rule of thumb: always work at a pH that is either 2 units above the protein's pI when using anion exchange or 2 units below the pI when using cation exchange. Of course, check to see if your protein can tolerate such a change in pH, but because there is always a choice of either cation or anion exchange, a suitable pH is generally available.

Choose the appropriate buffer and counter ion for the separation. Almost always sodium chloride is the appropriate choice for a counter ion. (On occasion, I use ammonium acetate, which is a volatile salt that buffers at either pH 4.5 or pH 9, so that I can perform electrospray mass spectrometry immediately on the fractions.) The choice of buffer is critical to maintaining the pH desired in the separation. Be sure to match the buffer's pK_a with the pH of the separation. Select a buffer that if charged will not stick to the column, e.g. use phosphate on cation exchange columns and TRIS-Cl on anion exchange columns. Remember anything charged will stick such as EDTA. Exercise caution when interpreting chromatograms, because a peak may really be EDTA or a buffer component and not protein

Choose a column volume as recommended by the manufacturer, keeping in mind that more capacity is better, and other things in the cell lysate will stick to your column such as cellular DNA. I use a column between 100 and 200 ml (50 mm diameter) for 0.1 to 2 gram scale separations. Two buffers will be required for the gradient. Buffer A is 20 to 50 mM of the buffer adjusted to the appropriate pH. Buffer B is Buffer A plus 1 M of the counter ion salt. Pre-equilibrate the column with two column volumes (CVs) of Buffer B, then five CVs of Buffer A. Adjust the pH of your sample to match that of Buffer A. Load the sample. Wash the column with 3 CVs of Buffer A and elute with a linear gradient of 0 to 100% Buffer B. The pace of the gradient depends on the degree of separation desired. Slower gradients separate better but broaden peaks more. Usually 10 CVs is an adequate gradient volume with a flow speed appropriate to the resin's specification sheet. Fractionate into tubes $\sim 1/50$ to $\sim 1/100$ of the gradient volume. Monitor with OD_{280nm} . As for any protein purification step, save everything until your

protein has been located. If there are many peaks on the chromatogram, SDS-PAGE is generally the method of choice for locating your sample.

A2.1.5 Six-His affinity chromatography

Some proteins are expressed with a small artificial 6-His tag on either the N- or C-terminus. These proteins are easily purified on metal chelating sephadex resins. These resins bind a divalent metal ion, like Ni^{2+} . The poly-histidine tag then coordinates onto the Ni^{2+} thus forming a specific interaction. This method is best employed in high salt (usually 2M sodium chloride) to reduce unwanted protein interactions from cellular proteins. The pH is crucial, and generally Buffer A is 20 mM phosphate and 2 M NaCl, pH 7.0. Buffer B is Buffer A plus 1 M imidazole. The imidazole competes for the metal ion binding site and elutes the protein selectively from the column. The column should be eluted after thoroughly washing with Buffer A to remove non-specific proteins that have loosely bound to the column. A gradient is preferable over 10 CVs. While the 6-His tag is popular for its convenience, the addition of the tag may introduce potential complications as it alters either the N- or C-terminus, and thus may consequently perturb the stability or folding pathway of the protein.

A2.1.6 Hydrophobic interaction chromatography

Follow many of the same guidelines for ion exchange; however, this resin is composed of hydrocarbon ligands, such as phenyl-sepharose. I use a buffer of acetic acid at pH 4 (0.1 % w/v). The buffers, however, are now set up exactly in reverse to a typical ion exchange run. Buffer B is the 0.1% acetic acid solution, pH 4, and Buffer A is Buffer

B plus 2 M ammonium sulfate. The principle of the separation is that the high salt conditions make hydrophobic residues on the surface of the protein stick to the phenyl groups on the column. Add 2 M ammonium sulfate to your protein sample and adjust the pH to 4. Load sample, wash with 2 CVs of Buffer A, and elute with a simple gradient to 100% Buffer B over 10 CVs. Note that this method is non-denaturing in contrast to HPLC below. Other neutral pH buffers, such as TRIS-Cl, also work well.

A2.1.7 HPLC

High performance liquid chromatography (HPLC) is the other workhorse of protein purification. If only one step is required, it is probably a single HPLC separation. When combined with ion exchange, this method is extremely powerful. I like to make HPLC the second step of a two step protocol, because the buffer, typically 0.1% trifluoroacetic acid, TFA, in acetonitrile and water, is highly volatile and excellent for freeze-drying. The major drawbacks of the method are the expense of the columns, their low binding capacity, the use of organic solvents, and the denaturing aspects of the separation. The method is excellent for protein folding or small globular proteins, since these proteins can be rapidly renatured following the separation (see below). Always acidify your sample with acetic acid or TFA and filter (0.2 μ m) before applying to the column. Buffer A is 0.1% TFA in water and Buffer B is 0.1% TFA in acetonitrile. I run gradients of 5% to 95% over 1 hour at 10 ml/min on a preparative HPLC column (22 mm diameter).

A2.1.8 Dialysis

This step is frequently performed to remove salts from protein solutions. Often prior to freeze-drying, extensive dialysis must be performed, because any residual non-volatile salt will cause thawing of your sample during freeze-drying. Usually three changes of water or dilute ammonium bicarbonate buffer eliminate the non-volatile salt concentration of your sample adequately enough for freeze-drying.

A2.1.9 Freeze-drying and storage

Dialyze the sample if it contains non-volatile salts. HPLC fractions can be used directly. Freeze the sample at -80°C or in liquid nitrogen in 50 ml disposable tubes. Remove cap from frozen tube, and fix a Kimwipe on top with a rubber band. Place on lyophilizer in bell jar. The sample is completely dry when it is light and fluffy. Poor lyophilization occurs when a sample has too much non-volatile salt or the vacuum is not sufficient. Freeze-drying is the most efficient way to concentrate protein and allows a protein to be stored indefinitely with minimal sample degradation. I store protein samples at -20°C in sealed 50 ml tubes.

A2.1.10 Reconstitution of freeze-dried proteins

While most proteins dissolve readily when lyophilized from water based purification steps, certain β -sheet containing proteins, like ubiquitin, do not readily dissolve in water after lyophilization from denaturing HPLC buffers that contain a high percentage of acetonitrile. The reasons for this are unclear but may be due to the entanglement of proteins into aggregates and the slower folding rates of β -sheet proteins. However, this situation can be rectified by dissolving the lyophilized powder in concentrated GdmCl buffer (6 to 7 M). If GdmCl must be removed, then dialyze against water.

Table A2.1 Purification Schemes

Protein	Purification Schemes ^a
Arc	recomb.; 6-His affinity; HPLC; freeze-dry
CI2	recomb.; anion (TRIS-Cl/NaCl/pH 7.5); HPLC; freeze-dry
CTACP	recomb.; cation (phosphate/NaCl/pH 7); hydrophobic; dialysis; freeze-dry
Cyt <i>c</i>	recomb.; cation (phosphate/NaCl/pH 7); HPLC; freeze-dry
GCN4 dimeric	synth.; HPLC; freeze-dry; HPLC; freeze-dry
GCN4 cross-linked	synth.; reduce ^b ; HPLC; freeze-dry; oxidize ^b ; HPLC; freeze-dry
IM7	recomb.; anion (TRIS-Cl/NaCl/pH 7.5); HPLC; freeze-dry
λ	recomb.; acid; HPLC; freeze-dry
SH3	recomb.; anion (TRIS-Cl/NaCl/pH 7.5); HPLC; freeze-dry
ProtG	recomb.; 6-His affinity; HPLC; freeze-dry
ProtL	recomb.; 6-His affinity; HPLC; freeze-dry
Ub	recomb.; acid; cation (acetate/ammonium acetate/pH 5); HPLC; freeze-dry

^aPlease refer to section A2.1 for exact protocols. Purification scheme abbreviations:

recomb., recombinant protein expression; synth., chemical peptide synthesis; acid, acetic acid precipitation; anion, anion exchange chromatography (buffer/counter ion/pH); cation, cation exchange chromatography (buffer/counter ion/pH). All buffers, which are not listed, are standard and explained in detail in section A2.1. All counter ion concentrations are 1 M.

^b“Reduce” and “oxidize” refer to disulfide chemistry procedures which form or break a Cys-Cys cross-link on GCN4. Reduction requires TCEP reductant at pH 7 (20 equivalents), whereas oxidation requires bubbling air through the solution at pH 8 for 1 hour.

Table A2.2 Protein Sequences

Protein ^a	Mutation ^b	Source ^c	Length (res.) ^d	Theor. Mol. Wt. (Da)	Act. Mol. Wt. (Da)	pI	Protein Sequence ^b
Arc	6-His tag ^f	recomb.g	64 ⁱ	7688 ⁱ	7690	9.99	MKMSKMPQF NLRWPREVLD LVRKVAEENG RSVNSEIYQR VMESEFKKEGR IGAHHHHHHK NQHE
Arc ^{MYL}	6-His tag ^f	recomb.g	64 ⁱ	7654 ⁱ	7655	9.52	MKMSKMPQF NLRWPREVLD LVRKVAEENG MSVNSYIYQL VMESEFKKEGR IGAHHHHHHK NQHE
CI2		recomb.	64	7292	7294	4.79	MKTEWPELVG KSVEEAKKVI LQDKPEAEII VLPVGTIVTM EYRIDRVRLF VDDLDNIAQV PRVG
CTACP		recomb.	99	11262	11131 ^e	9.30	MAEGNTLISV DYEIFGKVQG VFFRKHTQAE GKKLGLVGWV QNTDRGTVQG QLQGPISKVR HMQEWLETRG SPKSHIDKAN FNNEKVLKLL DYSDFQIVK
Cyt <i>c</i>		Sigma	104	12361 ^{m,n}	N.D.	9.59	GDVEKGKKIF VQKCAQCHTV EKGKHKHTGP NLHGLFGRKT GQAPGFTYTD ANKNKGITWK EETLMEYLEN PKYIIPGTM IFAGIKKKTE REDLIAYLKK ATNE
Cyt <i>c</i> ^{pWT j}	H26N, H33N	recomb. ^h	104	12273 ⁿ	N.D.	9.59	GDVEKGKKIF VQKCAQCHTV EKGKHKHTGP NINGLFGTKT GQAPGFTYTD ANKNKGITWK EETLMEYLEN PKYIIPGTM IFAGIKKKTE REDLIAYLKK ATNE
GCN4 ^d _{D7A}	D7A, W17W	synth.	36 ⁱ	4235 ^{i,m,o}	4235	9.30	CGGRMKQLEA KVEELLSSKNW HLENEVARLK KLUGER
GCN4 ^{Nx} _{D7A} ^k	D7A, W17W, N- term. cross-link ^k	synth.	72	8470 ^{m,o}	8470	9.30	CGGRMKQLEA KVEELLSSKNW HLENEVARLK KLUGER
GCN4 ^{d,A24G} _{biH3,7}	K3H, D7H, Y17W, A24G	synth.	35 ⁱ	4169 ^{i,m,o}	4168	9.16	GGRMHQLEHK VEELLSSKNWN LENEVGRLLK LVGER

Table A2.2 Protein Sequences (continued)

Protein ^a	Mutation ^b	Source ^c	Length (res.) ^d	Theor. Mol. Wt. (Da)	Act. Mol. Wt. (Da)	pI	Protein Sequence ^b
GCN4 ^{Nx-A24G} _{biH3,7} k	K3H, D7H, Y17W, A24G, N- term. cross-link k	synth.	72	8545m,o	8545	9.16	CGGRMHQLEH KVEELLSKNW NLENEVGRLK KLVGER
GCN4 ^d _{biH3,7}	K3H, D7H, Y17W	synth.	35 ⁱ	4183i,m,o	4185	9.16	RMHQLEHKVE ELLSKNWNLE NEVARLKKLV GERGC
GCN4 ^{Cx} _{biH3,7} k	K3H, D7H, Y17W C-term. cross-link k	synth.	72	8573m,o	8573	9.16	RMHQLEHKVE ELLSKNWNLE NEVARLKKLV GERGC
IM7		recomb.	87	9896	9896	4.46	MELKNSISDY TEAEFVQLLK EIEKENVAAT DDVLDVLLLEH FVKITEHPDG TDLIYYPSDN RDDSPEGIVK EIKEWRAANG KPGFKQG
λ^{GG}	residues 6-85 ^l	recomb.	81	8879	8747 ^e	8.04	MSLTQEQLD ARRLKAIWEK KKNELGLSQE SVADKMGMGQ SGVGALFNGI NALNAYNAAL LAKILKVSVE EFSPSIAREI R
λ^{AA}	G46A, G48A, residues 6-85 ^l	recomb.	81	8907	8777 ^e	8.04	MSLTQEQLD ARRLKAIWEK KKNELGLSQE SVADKMGMGQ SAVAALFNGI NALNAYNAAL LAKILKVSVE EFSPSIAREI R
SH3		recomb.	66	7351	7220 ^e	4.83	MGSNMGGVTT FVALYDYESR TETDLSFKKG ERLQIVNNTE GDWWLAHSLT TGRTGYIPSN YVAPSD
ProtG	6-His tag ^f	recomb.	65	7337	7337	5.56	MHHHHHAMD TYKLJLNGKT LKGETTTEAV DAATAEKVFK QYANDNGVDG EWTYDDATKT FTVTE
ProtL	Y47W, 6-His tag ^f	recomb.	72	8041	8043	6.01	MHHHHHAME EVTIKANLIF ANGSTQTAEF KGTFEKATSE AYAYADTLKK DNGEWTVDVA DKGYTLNKF AG
Ub ^{WT}		recomb.	76	8566	8565	6.56	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLV LRLRG

Table A2.2 Protein Sequences (continued)

Protein ^a	Mutation ^b	Source ^c	Length (res.) ^d	Theor. Mol. Wt. (Da)	Act. Mol. Wt. (Da)	pI	Protein Sequence ^b
Ub	F45W	recomb.	76	8605	8604	6.56	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLHLV LRLRG
Ub ^{A28G}	F45W, A28G	recomb.	76	8591	8590	6.56	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLHLV LRLRG
Ub ^{pWT}	F45W, H68N	recomb.	76	8582	8582	6.21	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLNLV LRLRG
Ub ^{H1}	F45W, H68N, E24H, A28H	recomb.	76	8656	8655	8.08	MQIFVKTLTG KTITLEVEPS DTIENVKHKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLNLV LRLRG
Ub ^{H2}	F45W, H68N, A28H, D32H	recomb.	76	8670	8669	8.08	MQIFVKTLTG KTITLEVEPS DTIENVKHKI QHKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLNLV LRLRG
Ub ^{B1}	F45W, K6H	recomb.	76	8614	8612	6.06	MQIFVHTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLHLV LRLRG
Ub ^{B2}	F45W, H68N, Q2H, E16H	recomb.	76	8599	8597	8.08	MHIFVKTLTG KTITLHVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKESTLNLV LRLRG
Ub ^{B3}	F45W, H68N, Q2H, E64H	recomb.	76	8599	8598	8.08	MHIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGRTLSDYN IQKHSTLNLV LRLRG
Ub ^{T1}	F45W, H68N, R54H, D58H	recomb.	76	8585	N.D.	6.72	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIWAGKQL EDGHTLSHYN IQKESTLNLV LRLRG

^aAdditional information on these proteins is provided in Chs. 3 and 4 and App. 1.

Table A2.2 Protein Sequences (continued)

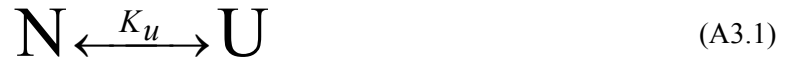
- ^bSequence numbering conventions of site directed mutants vary according to each system.
- ^cSource: peptide synthesis (synth.), recombinant protein expression (recomb.), or purchased from a vendor (e.g. Sigma).
- ^dLength of expected translation product; actual length could vary with co-translational processing of the N-terminal methionine.
- ^eMol. wt. discrepancy of ~131 Da is most likely due to co-translational processing of the N-terminal methionine.
- ^fSix-His purification tags vary in sequence and position.
- ^gRecombinant protein provided by R. Sauer (MIT).
- ^hRecombinant protein provided by S. W. Englander (U. of Pennsylvania).
- ⁱLengths and molecular weights of dimeric proteins are reported for a single monomer.
- ^jPseudo-wild type version of cyt *c*, which prevents His-heme mis-ligation, is produced recombinantly.
- ^kIncludes either a Cys-Gly-Gly N-terminal or Gly-Gly-Cys C-terminal linker that cross-links when oxidized into a disulfide bond.
- ^lSequence includes only residues 6-85 which makes a monomeric version of λ .
- ^mMol. wt. and sequence include an N-terminal acetylation.
- ⁿMol. wt. includes a heme prosthetic group.
- ^oMol. wt. and sequence include a C-terminal amidation.

Appendix 3

Two-State Equilibrium and Kinetic Analysis

A3.1 Monomeric systems

A3.1.1 Two-state equilibrium stability derivation



$$K_u = \frac{[\text{U}]}{[\text{N}]} \quad (\text{A3.2})$$

$$[\text{U}] = C_T \times f_U \quad [\text{N}] = C_T \times f_N \quad (\text{A3.3})$$

$$f_N = \frac{\Theta}{N_{\text{Baseline}} - U_{\text{Baseline}}} \quad (\text{A3.4})$$

$$f_U = 1 - f_N \quad (\text{A3.5})$$

$$K_u = \frac{f_U}{f_N} = \frac{1 - f_N}{f_N} = \frac{1}{f_N} - 1 \quad (\text{A3.6})$$

$$\Theta = \frac{N_{\text{Baseline}} - U_{\text{Baseline}}}{K_u + 1} \quad (\text{A3.7})$$

$$\Delta G_u = -RT \ln K_u = \Delta G^o - m^o \cdot [\text{Den}] \quad (\text{A3.8})$$

$$K_u = \exp\left(\frac{-\Delta G^o + m^o \cdot [\text{Den}]}{RT}\right) \quad (\text{A3.9})$$

$$\Theta = \frac{N_{\text{Baseline}} - U_{\text{Baseline}}}{\exp\left(\frac{-\Delta G^o + m^o \cdot [\text{Den}]}{RT}\right) + 1} \quad (\text{A3.10})$$

Standard two-state equilibrium denaturant melt experiments usually require fitting an observed signal, Θ , using sloping baselines. Notice that for monomeric systems the concentration term, C_T , cancels (A3.3-4), and the stability is only reliant of the fraction

native, f_N , and unfolded, f_U . The denaturant dependence of the protein's stability is the m -value, which is in units of $\text{kcal mol}^{-1} \text{M}^{-1}$ (A3.8). This value corresponds to the amount of denaturant sensitive surface area buried in the native state relative to the unfolded state. The general equation (A3.10) fits the observed signal as a function of denaturant, $[Den]$, yielding the stability in the absence of denaturant, ΔG^o , and the m -value.

A3.1.2 Two-state kinetics derivation



$$\frac{\partial U}{\partial t} = k_u[N] - k_f[U] = -\lambda[U] \quad (\text{A3.12})$$

$$\frac{\partial N}{\partial t} = -k_u[N] + k_f[U] = -\lambda[N] \quad (\text{A3.13})$$

$$\begin{vmatrix} \lambda - k_f & k_u \\ k_f & \lambda - k_u \end{vmatrix} \quad (\text{A3.14})$$

$$\lambda^2 - \lambda(k_f + k_u) = 0 \quad (\text{A3.15})$$

$$\lambda = k_f + k_u \quad (\text{A3.16})$$

$$\Delta G_u^\ddagger([Den]) = \Delta G_u^{\ddagger, H_2O} + m_u[Den] + Const \quad (\text{A3.17})$$

$$\Delta G_f^\ddagger([Den]) = \Delta G_f^{\ddagger, H_2O} + m_f[Den] + Const \quad (\text{A3.18})$$

$$k_u([Den]) = \exp(\Delta G_u^\ddagger([Den]) / RT) \quad (\text{A3.19})$$

$$k_f([Den]) = \exp(\Delta G_f^\ddagger([Den]) / RT) \quad (\text{A3.20})$$

$$m^o = m_u - m_f \quad (\text{A3.21})$$

$$\Delta G_{eq}^o = \Delta G_u^\ddagger - \Delta G_f^\ddagger \quad (\text{A3.22})$$

$$RT \ln \lambda = RT \ln(k_f([Den]) + k_u([Den])) \quad (\text{A3.23})$$

The differential equations (A3.12-13) for the two-state folding reaction (A3.11) are solved here by the matrix method (A3.14), which obtains the solution for the

observed rate, λ , from the eigenvalues: the single non-zero solution is the sum of the folding, k_f , and unfolding rates, k_u (A3.16). The activation energies of these rate constants (A3.17-18) are linearly proportional to the denaturant concentration. Confirmation of two-state folding is given by the equality of the sums of the kinetically determined m -values and activation energies with respective equilibrium m -value and stability measurements (A3.21-22). Finally, the chevron equation is a combination of the solution of the matrix and the denaturant dependencies of the folding and unfolding rates (A3.23; see also Sec. A3.3).

A3.2 Dimeric systems

A3.2.1 Two-state equilibrium stability derivation



$$K_u = \frac{[U]^2}{[N]} \quad (\text{A3.25})$$

$$[U] = C_T \times f_U \quad [N] = \frac{1}{2} C_T \times f_N \quad (\text{A3.26})$$

$$f_U = \frac{N_{\text{Baseline}} - \Theta}{N_{\text{Baseline}} - U_{\text{Baseline}}} \quad f_N = \frac{\Theta - U_{\text{Baseline}}}{N_{\text{Baseline}} - U_{\text{Baseline}}} \quad (\text{A3.27})$$

$$K_u = \frac{f_U^2 \times C_T^2}{\frac{1}{2} f_N \times C_T} = \frac{(N_{\text{Baseline}} - \Theta)^2}{\frac{1}{2} (N_{\text{Baseline}} - U_{\text{Baseline}}) (\Theta - U_{\text{Baseline}})} C_T \quad (\text{A3.28})$$

$$N \equiv N_{\text{Baseline}} \quad U \equiv U_{\text{Baseline}} \quad B \equiv N_{\text{Baseline}} - U_{\text{Baseline}} \quad (\text{A3.29})$$

$$\frac{BK_u}{2C_T} = \frac{(N - \Theta)^2}{\Theta - U} \quad Z \equiv \frac{BK_u}{2C_T} \quad (\text{A3.30})$$

$$0 = \Theta^2 - (2N + Z)\Theta + (ZU + N^2) \quad (\text{A3.31})$$

$$\Theta = N + \frac{Z}{2} \pm \frac{1}{2} \sqrt{(2N + Z)^2 - 4(ZU + N^2)} \quad (\text{A3.32})$$

$$\Theta = N + \frac{Z}{2} \pm \frac{1}{2} \sqrt{Z^2 + 4ZB} \quad (\text{A3.33})$$

$$\Theta = N + \frac{BK_u}{4C_T} \pm \frac{1}{2} \sqrt{\frac{B^2 K_u^2}{4C_T^2} + \frac{2B^2 K_u}{C_T}} \quad (\text{A3.34})$$

$$\Theta = N + \frac{BK_u}{4C_T} \pm \frac{B}{4C_T} \sqrt{K_u^2 + 8K_u C_T} \quad (\text{A3.35})$$

$$\Theta = N + \frac{B}{4C_T} \left(K_u - \sqrt{K_u^2 + 8K_u C_T} \right) \quad (\text{A3.36})$$

$$\Theta = N + \frac{BK_u}{4C_T} \left(1 - \sqrt{\frac{8C_T}{K_u} + 1} \right) \quad (\text{A3.37})$$

$$\Delta G_u = -RT \ln K_u = \Delta G^o - m \times [\text{Den}] \quad (\text{A3.38})$$

$$K_u = \exp \left(\frac{-\Delta G^o + m \times [\text{Den}]}{RT} \right) \quad (\text{A3.39})$$

While analogous to monomeric systems in terms of the denaturant dependence (A3.38), the concentration term, C_T , also determines the apparent stability of a dimeric system (A3.28). Generally, for dimeric systems the stability at 1 M standard state is reported, combining A3.37-39; however, a more realistic quantity may be the dimeric protein's stability at 1 μM , since the former quantity is physically impossible to achieve in experiment.

A3.2.2 Two-state kinetics equations



$$\Delta G_f^\ddagger([\text{GdmHCl}]) = -RT \ln k_f^{H_2O} - m_f [\text{Den}] - RT \ln C_T + \text{Constant} \quad (\text{A3.41})$$

$$\Delta G_u^\ddagger([\text{GdmHCl}]) = -RT \ln 2k_u^{H_2O} - m_u [\text{Den}] + \text{Constant} \quad (\text{A3.42})$$

Instead of rigorously deriving the kinetics from the differential equations for a dimeric folding reaction (A3.40) in the presence of a denaturant, I have simply provided the denaturant dependencies of the folding and unfolding arms. Notice that only the folding rate is dependent on the concentration of unfolded monomer, or C_T (A3.41). The analytical solution to the dimerization kinetics yields a constant factor of two, which is factored with the unfolding rate constant, $k_u^{H_2O}$ (A3.42). The m -values are the same as defined for the monomeric derivation (Sec. A3.1.2); however, these dimeric forms can only be applied under strong folding and unfolding conditions. When folding or unfolding reactions are monitored in the transition region near the equilibrium midpoint, more complicated analysis is required. Kinetics work presented in this thesis, therefore, was carried out under strong folding and unfolding conditions to avoid these complications. Finally, both ϕ - and ψ -analysis may be applied to dimeric folding systems (see Sec. A3.3 and Ch. 4).

A3.3 ϕ -analysis

ϕ -analysis (Matthews, 1987) is an application of Brønsted analysis whereby the degree to which an interaction is present at the rate-limiting step for the protein folding reaction is assessed.

$$\Delta\Delta G_{eq} = K_{eq} / K_{eq}^* \quad (\text{A3.43})$$

$$\Delta\Delta G_f^\ddagger = k_f / k_f^* \quad (\text{A3.44})$$

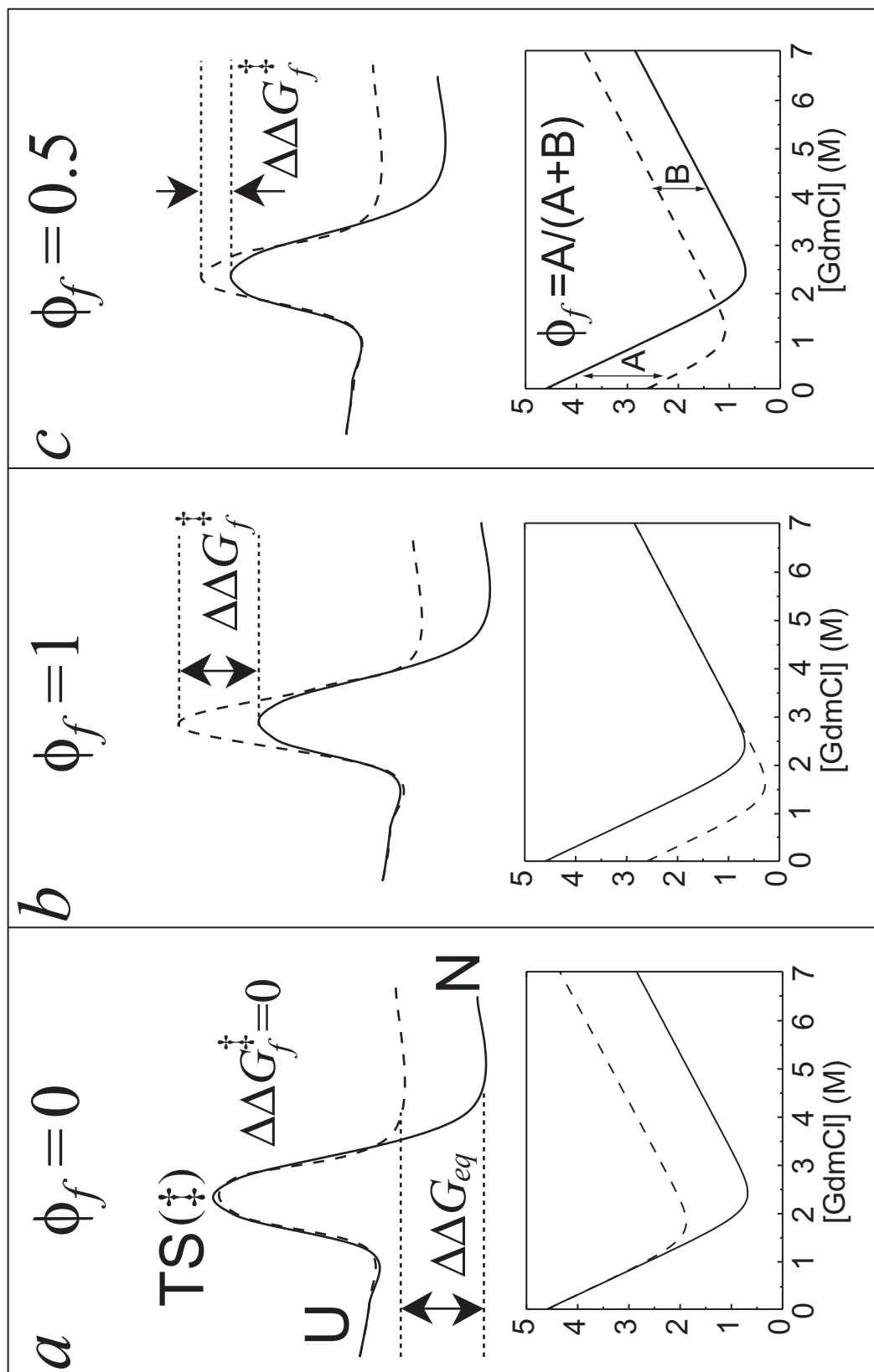
$$\Delta\Delta G_f^\ddagger = \phi \times \Delta\Delta G_{eq} \quad (\text{A3.45})$$

Generally, an amino acid is mutated (where the rate and equilibrium constants above are designated with an *). However, any perturbation such as isotopic substitution (Ch. 3), addition of metal ions (Ch. 4), addition of a ligand or co-solvent, or change in pH could be applied in ϕ -analysis (A3.45). If the perturbation in terms of free energy, or $\Delta\Delta G_{eq}$, (A3.43) solely affects the folding activation energy then the site of perturbation is said to have a ϕ -value of one. If, however, a $\Delta\Delta G_{eq}$ for a perturbation is entirely accounted for in the unfolding activation energy, then the ϕ -value is zero.

When the $\Delta\Delta G_{eq}$ for a perturbation is realized in both the folding and unfolding activation energies, then the ϕ -value is fractional and has an inherently complex interpretation, because ϕ -values are a slope from a single parameter fit to only two data points (A3.45). Thus a partial ϕ -value may either indicate (i) that multiple folding pathways exist where the site which was perturbed stabilizes or destabilizes one pathway relative to the other or (ii) that the site which was perturbed is partially formed in a singular transition state. Of course, these two possibilities may be combined, further emphasizing the ambiguity in partial ϕ -values.

Often the concept of ϕ -analysis and this particular ambiguity confounds theorists who envision multiple rate-limiting steps for multiple folding routes or that the transition state is rather a broad and diffuse species which is not readily defined. I have resolved this inherent ambiguity in Chapter 4 using ψ -analysis, where gradual increases in the perturbation free energy are available by titration of a metal ion. Thus two parameter fits of the folding activation energy profile are attained. The first parameter, the f -value, indicates the degree to which an interaction is present in the transition state, and the second parameter, the ρ -value, indicates the level of pathway heterogeneity (see Ch. 4).

Figure A3.1 ϕ -values and chevron plots. ϕ -analysis is the application of reaction rate theory to chevron plots (A3.23): for a two-state system, the effects of an energetic perturbation to the protein's equilibrium stability is manifested in either the folding or unfolding activation energies or both. In such analyses, it is generally assumed that the mutation only changes the stability of the native state. As defined, $\phi = \Delta\Delta G_f^\ddagger / \Delta\Delta G_{eq}$ (A3.45). Thus a mutation that only affects the folding activation energy (**a**) will only alter the protein's folding rate. The interaction is, therefore, fully formed and native-like in the rate-limiting step, or transition state (TS), and the ϕ -value is unity. However, when only the unfolding arm of the chevron is altered (**b**), then the ϕ -value is zero, and the interaction is completely lacking in the TS. Nonetheless, most mutational ϕ -values are fractional (**c**). In such cases, ϕ -analysis is ambiguous. Either the interaction may be partially formed in the TS, a fraction of the population may form the interaction fully, or a combination of these two possibilities. Only more sophisticated analysis, such as metal binding ψ -analysis, can formally distinguish these possibilities (see Ch. 4).



Appendix 4

Optimizing the Stopped-flow

A4.1 Hardware

The appendix that follows discusses specific optimizations made to Biologic's SFM stopped-flow. Much of this information is meant to pose as a specific construction manual for similar modifications to Biologic stopped-flows as well as a reference source for ideas which improve other brands of stopped-flow instruments. The appendix is broken up into two major sections. First, the hardware section goes through the basics of interfacing a light source to the observation head, choice of a light source, and finally the photosensor module and its associated electronics. Second, I deal with the post-acquisition software filtering and other computer scripts that facilitate stopped-flow data acquisition, allowing for high data throughput.

A4.1.1 Redesigned observation head

Biologic manufactures their own fiber optic interface to their stopped-flow observation head; however, their design fails to completely illuminate the entire flow cell and requires additional focusing optics. The design schematized in Fig. A4.1 illustrates that a black delrin plastic housing can be built which holds the photosensor module (PMT) as well as an input light fiber optic. The design is flexible enough to allow for an additional absorbance fiber optic to be installed.

A4.1.2 Redesigned hard stop case

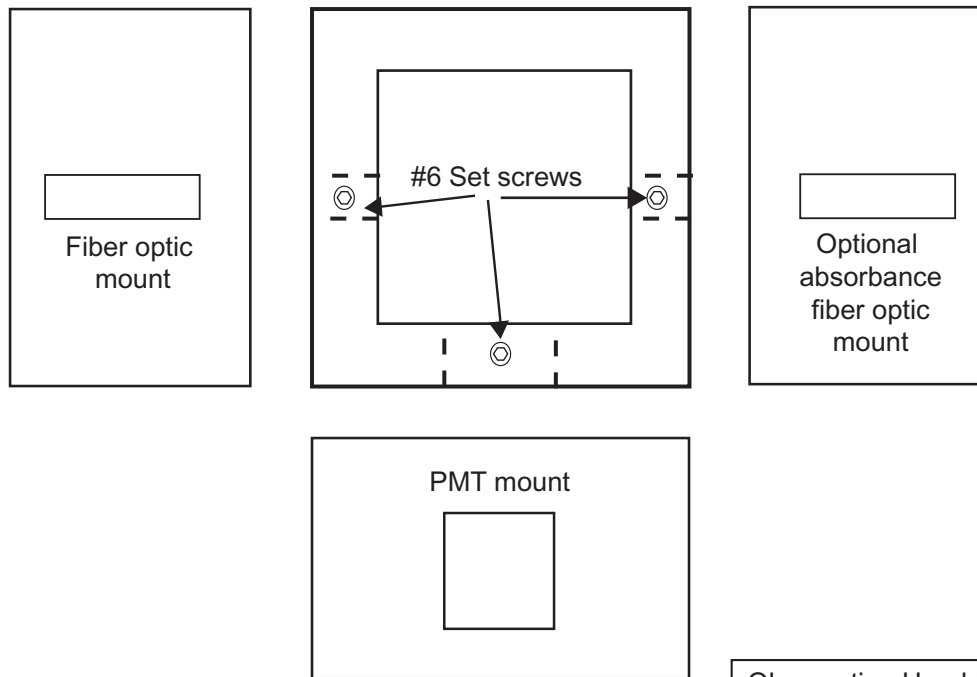
The electromechanical hard stop valve made by Parker-Hannifin allows the Biologic stopped-flow to rapidly cease flow. This component is essential to operation at fast flow speeds approaching 15 ml s^{-1} . However, as with most electromechanical components that fire repeatedly, this hard stop valve is prone to over-heating. Over-heating changes the character of the stop, affecting the quality of the data. I alleviate this problem by changing the manufacture's design by fixing an adjustable aluminum heat sink jacket around the valve (Fig. A4.2).

Another entirely different set of practical issues regarding the hard stop valve is the tension of an internal spring inside the valve. Usually a 0.050" set screw is located on top of the valve, which can be adjusted *slightly* to change the spring's tension. Generally, a stop is "too hard" when the spring is too tight and thus an early time glitch in the fluorescence trace is apparent. Fix by adjusting in small 1/8th to 1/4th of a turn increments. A similar problem also occurs when the spring is too relaxed and the stop is correspondingly "too soft" and has a late time artifact. Fix using similar 1/8th to 1/4th turn adjustments. Always test the valve after each incremental adjustment using an acquisition time of about 20 s to be able to witness both the early and late time artifacts.

The flow speed of the stopped-flow also affects the hard stop valve's calibration. When dramatically changing flow speeds expect that the stop can change and potentially the valve will require re-adjustment. This relationship to the flow speed demonstrates that the hard stop must oppose the momentum of the solution, which varies proportionally with flow speed.

Figure A4.1 Redesigned stopped-flow observation head. The observation head interface (*top*) is machined to fit the existing Biologic observation head (bottom left). The PMT light baffle (bottom right) reduces background levels of light scattering inside the observation head, improving signal-to-noise.

PMT/Fiber Optic Interface to SFM-4 Observation Head



Observation Head Interface
Black Delrin

Light baffle reduces background light scattering and interfaces PMT to flow cell

PMT Light Baffle
Anodized Aluminum

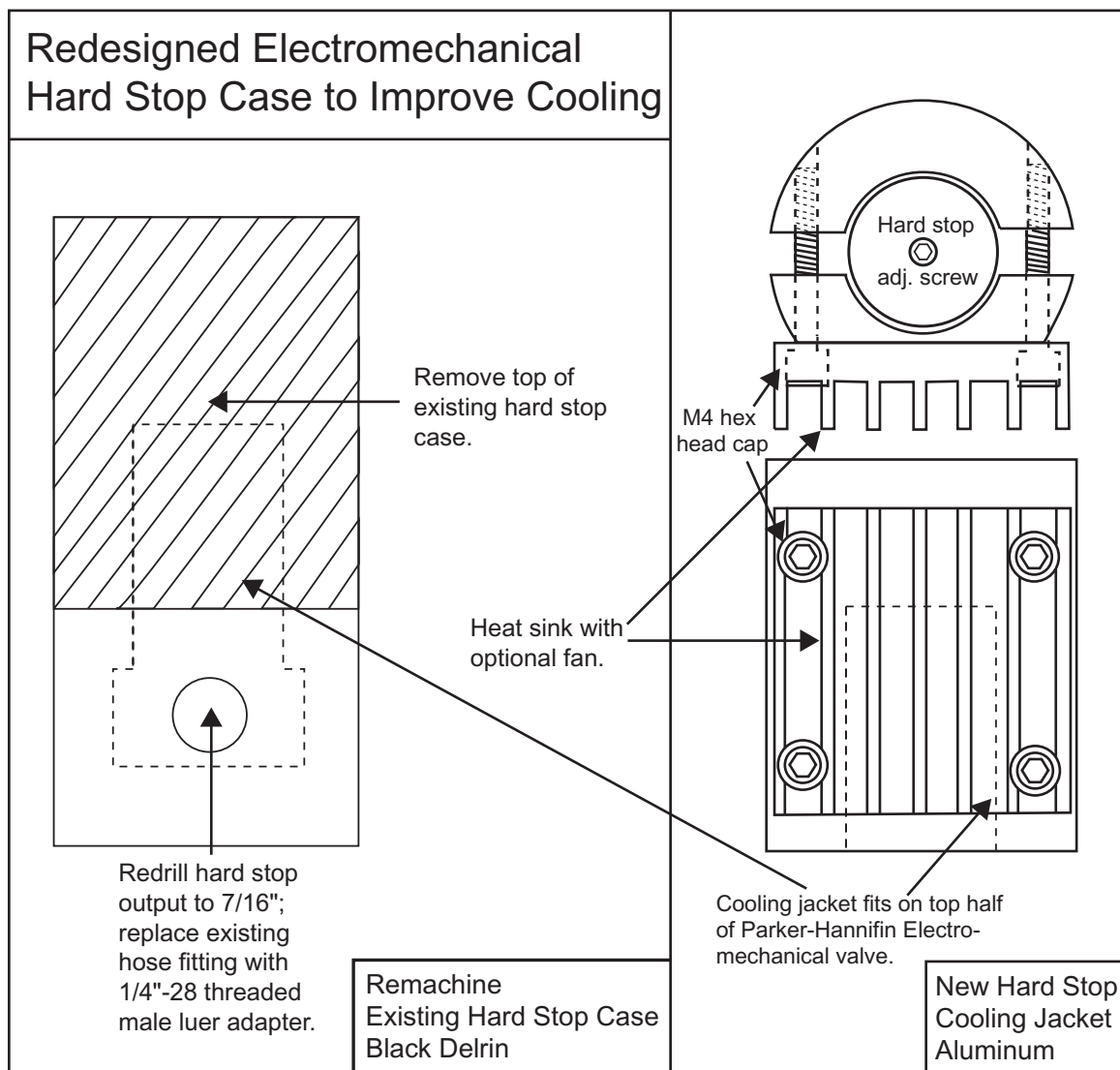
Theaded for Hard Stop Cap

Remove SS material for close approach of fiber optic to flow cell. Leave intact for PMT light baffle

Obs. head M4 mounting screw holes

Remachine Existing Observation Head Stainless Steel

Figure A4.2 Redesigned hard stop with heat sink and fan cooling. Note the position of the hard stop adjustment screw. This screw allows for the tension of the spring inside the valve to be tightened or relaxed, modifying the character of the stop.



A4.1.3 Optics

Probably the greatest single improvement to the stopped-flow's signal-to-noise was the light source. Increasing the light intensity at the excitation frequency in any fluorescence measurement improves the signal-to-noise, because the chromophore produces more signal and the high tension on the PMT can be lowered, reducing electronic noise. We chose the ellipsoidal light focusing optics of the PTI light source, because more solid angle of the lamp was focused than conventional lens focusing optics. We only used a single monochromator to further increase the light intensity.

I found that cooling the light source with -10 °C coolant (20% ethylene glycol w/v) improved lamp fluctuations and electronic noise in the igniter associated with overheating. The latter noise is serious and indicates poor cooling, and the character of electronic noise is “spiking” to ground at 60 Hz.

A4.1.3.1 Mercury versus xenon arc lamps

While the choice of the light source for fluorescence depends upon the chromophore, proteins containing tryptophan, unlike RNA and DNA, uniquely possess intrinsic fluorescence. The mercury arc lamp provides a bright band at 280 nm, which perfectly matches tryptophan's excitation wavelength. The mercury lamp outperforms the xenon lamp at 280 nm by 100-fold, especially when properly cooled (see section A4.1.3). Nonetheless, when relying on wavelengths other than 280 nm or the other available mercury lines, the smoother emission spectrum of a xenon lamp may be better suited.

A4.1.3.2 Reducing observation head background light scattering

A further improvement in signal-to-noise was made by creating a light baffle that interfaced the PMT with the emitted light from the flow cell. Since the excitation light scattering within the observation head contaminates the fluorescence signal, this baffle improves the data quality by about 5-fold (see Fig. A4.1).

A4.1.3.3 Fluorescence versus absorbance measurements

For fluorescence measurements, a glass cut-off filter is used to eliminate most of the light below 310 nm (for intrinsic Trp fluorescence). However, when conducting absorbance measurements, either a high quality notch filter or second monochromator is required. Essentially parasitic wavelengths of light contaminate the input spectrum, since it is a single monochromatic light source. Thus adding a second monochromator for the PMT at the output allows accurate absorbance measurements to be made.

A4.1.4 Electronics

Schematic diagrams are provided for the basic electronic interface between the PMT, its power supply, RC signal filtering, elliptical function filtering, signal monitoring and finally an interface to a computer through an A/D board.

A4.1.4.1 Hamamatsu photomultiplier tube

The Hamamatsu PMT was chosen because of the built in amplifier for the high tension as well as the on-board current to voltage conversion. The PMT responds well to the wavelengths utilized by intrinsic Trp fluorescence (300-400 nm); however, the PMT has poor response beyond 650 nm, which is not atypical of PMT's.

A4.1.4.2 Electronic interface to A/D board

The electronic interface houses all the interfacing power supply and signal filtering components in one location although a separate enclosure is used for the elliptical function filter, since it is sensitive to the AC power required for the power supply. The EMI surge suppressor reduces AC power line noise. A linear AC to DC converter is used to provide the necessary ± 12 V required for the PMT. I convert +12 V to +5V using a LM7805, providing power to a simple LCD display, which continuously monitors the signal and other helpful LED indicators. A parts lists is provided in Table A4.2.

A4.1.4.3 Signal filtering and over-sampling

Most equipment that measures a signal filters in some capacity the frequencies that are either too fast or too slow to be relevant to the acquisition frequency or the experimentally measured quantity of interest. Low-pass filters can be applied to remove the high frequency noise from the measurement. The simplest filter is the RC filter (Resistor Capacitor). These filters integrate the noise by filling the capacitor, according to the equation, $R \times C = \tau$, where τ is the time constant (in seconds) that defines the 3 dB cut-off frequency. Better filtering is provided by more complex filters (LRC) composed of inductors, L , which when connected in multiple sections, filter with a very sharp cut-off. By first passing through a 20 kHz elliptical function LRC filter (TTE, Inc.), the signal is then processed by a 20 kHz, 10kHz, 5 kHz, or 2 kHz RC section. The latter RC section is chosen depending on the acquisition frequency and the observed kinetic rate. Figure A4.7 demonstrates the results of the different filtering steps. Finally, the A/D board can over-sample a signal, allowing for boxcar averaging to further reduce noise. In the following section, I discuss a logarithmic filter step implemented prior to fitting the data.

Figure A4.3 Overall layout of the Biologic stopped-flow instrument

Figure A4.4 PMT power supply. The PMT power supply is a linear AC to DC power supply. The alternative, a switching power supply, is less well-regulated than a linear powers supply and contains more residual oscillation (or ripple) from the input AC.

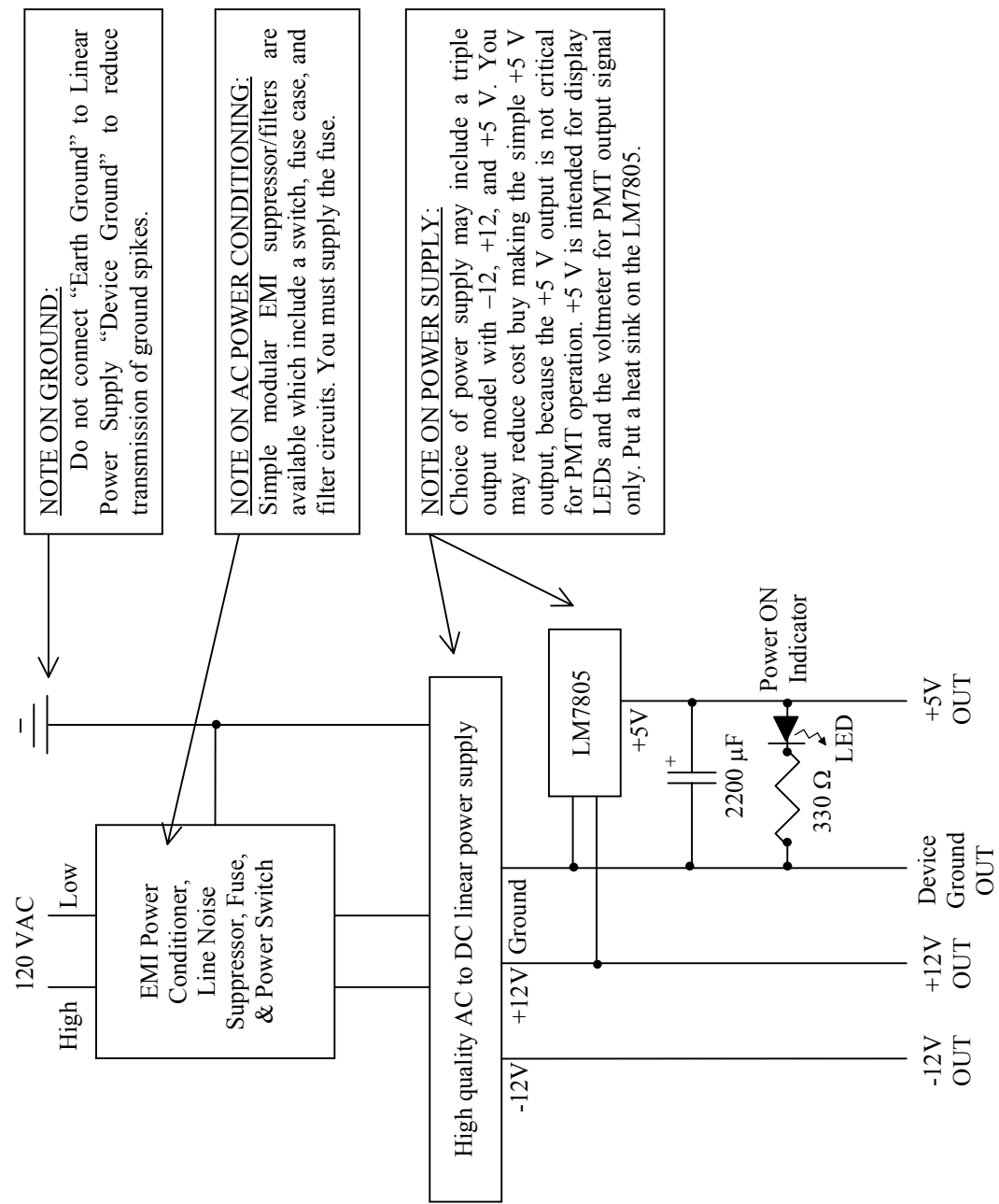


Figure A4.5 PMT control and signal filtering electronics. PMT power supply also contains PMT gain control, a LCD for signal display, RC filtering and a BNC breakout panel.

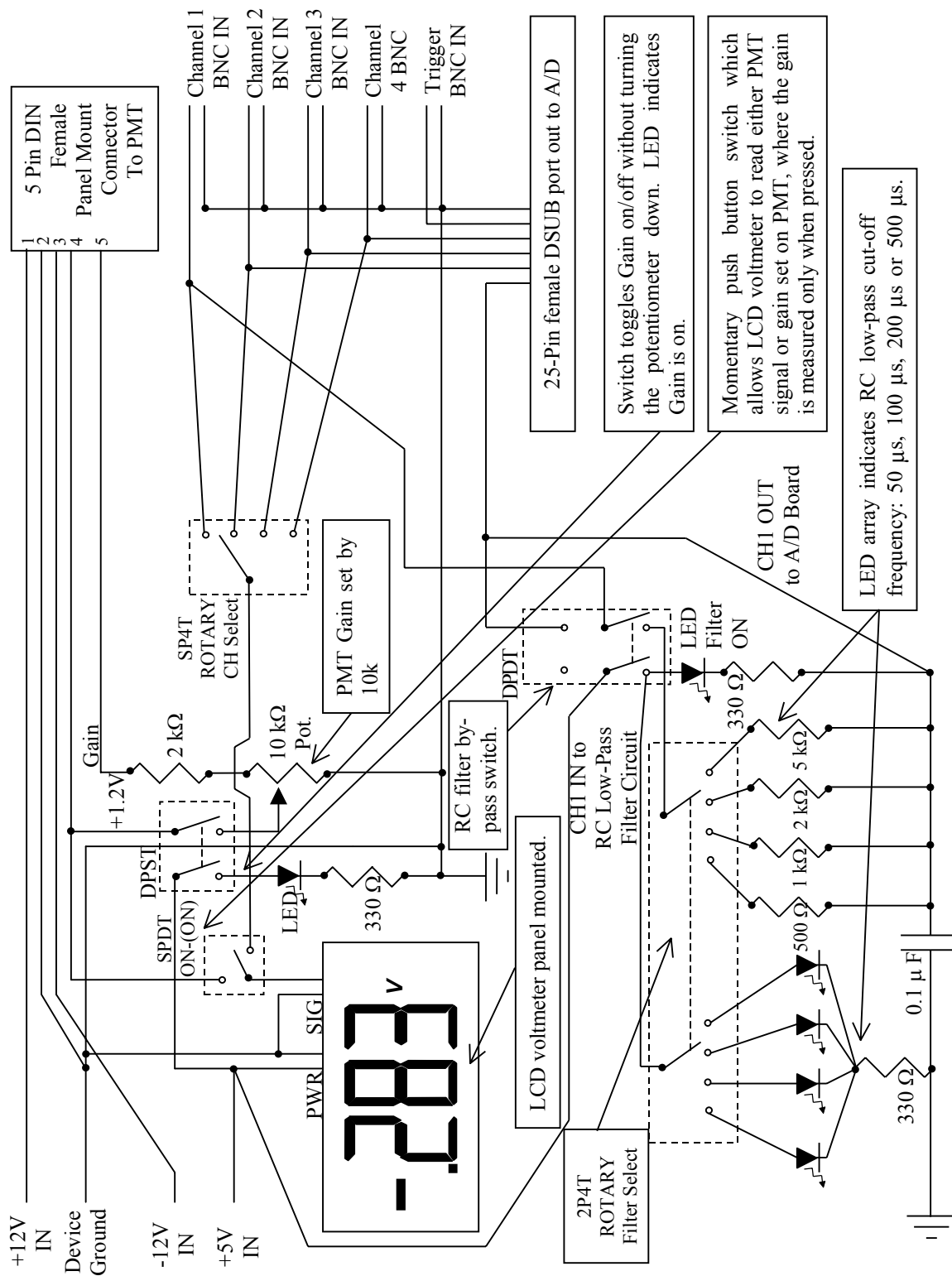


Figure A4.6 PMT signal pre-filtering with an LRC. The fluorescence signal is pre-filtered by the elliptical function, low pass filter (TTE, Inc.). This schematic allows for one of two frequencies to be selected or the pre-filter may be by-passed. This filter assembly is enclosed in a separate aluminum box to reduce the interference transmitted by the AC lines in the power supply enclosure.

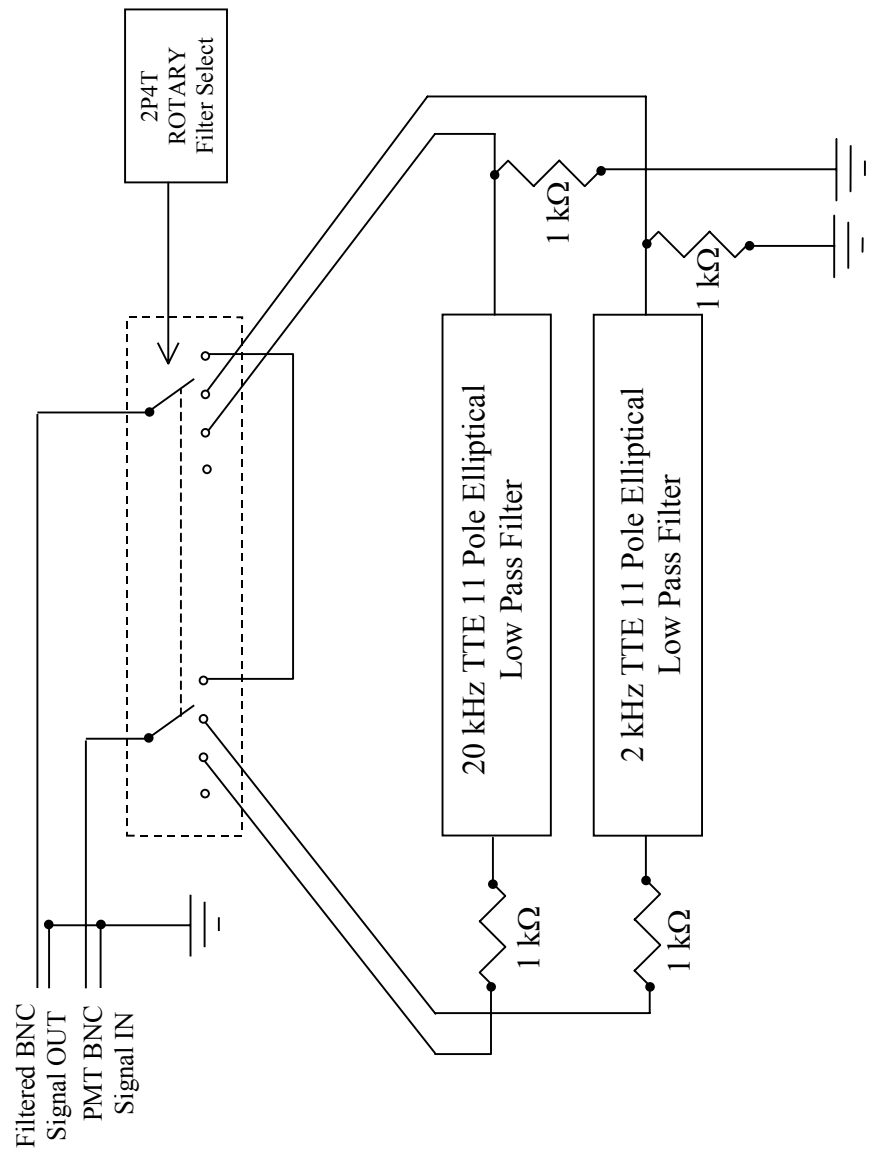
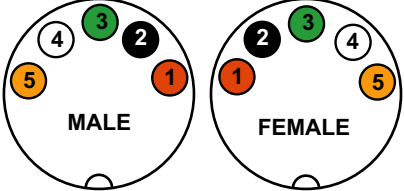
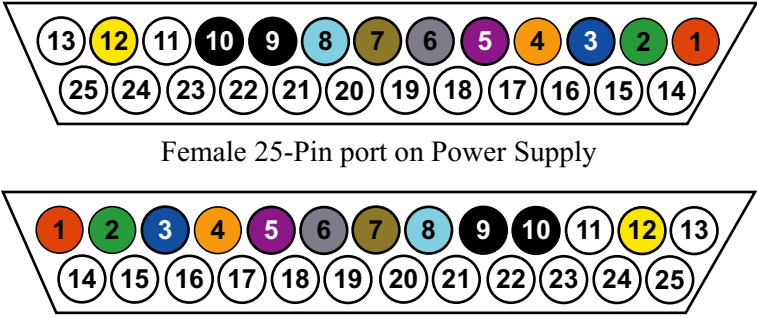
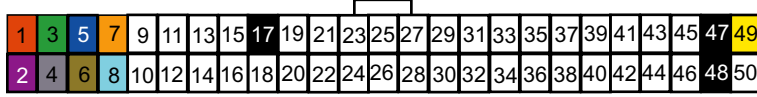


Table A4.1 Pinouts for cables created to interface the electronic components^a

Drawing	Pin	Color	Description
5-Pin DIN PMT Connector 	1	red	+12 V power
	2	black	Gnd
	3	green	-12 V power
	4	white	V_{control} / Gain
	5	orange	+1.2 V _{out}
25-Pin DSUB Power Supply to A/D Board Interface Connector 	1	red	CH1 (+)
	2	green	CH2 (+)
	3	blue	CH3 (+)
	4	orange	CH4 (+)
	5	purple	CH1 (-) ^b
	6	grey	CH2 (-) ^b
	7	brown	CH3 (-) ^b
	8	cyan	CH4 (-) ^b
	9	black	Gnd
	10	black	Gnd
	11		N.C.
	12	yellow	Trigger
50-Pin IDC female end of A/D Board Adapter Cable 	13-25		N.C.
	1	red	CH1 (+)
	2	purple	CH1 (-) ^b
	3	green	CH2 (+)
	4	grey	CH2 (-) ^b
	5	blue	CH3 (+)
	6	brown	CH3 (-) ^b
	7	orange	CH4 (+)
	8	cyan	CH4 (-) ^b
	9-16		N.C.
	17	black	Gnd
	18-46		N.C.
	47	black	Gnd
	48	black	Gnd
	49	yellow	Trigger ^c
	50		N.C.

^aAlternative pinouts may be used for the interface of the PMT to the power supply; the interface to the DT2801 A/D board must use the 50 pin connector. I choose a 25-pin DSUB to connect to the power supply box, because only 11 pins are really needed for acquisition. Thus, this table shows the connections required to make the 25-pin/50-pin adapter cable.

^bCH1 (-), CH2 (-), CH3 (-) and CH4 (-) are connected to ground inside the PMT power supply electronics case; they are essentially for true differential signal inputs for the 4 channels. They would pair with their respective (+) counterparts; however, a simplified design where these (-) inputs are grounded and signals are only connected to the (+) polarity is sufficient. In theory, these could be used to set an arbitrary offset.

^cA TTL pulse triggers the A/D board to start data acquisition.

Table A4.2 Complete parts list for stopped-flow modifications

Large Modular Components				
Qty.	Part #	Vendor	Stopped-Flow Component	Description
1		PTI	Light Source	100 Watt mercury arc lamp with ellipsoidal reflector
1	77539	Oriel	Light Source	Quartz fiber optic cable 3ft long (0.8×9.7 mm rectangular to 3.2 mm round connectors)
1	002-0508-900	Parker-Hannifin	Hard Stop	Replacement electromechanical hard stop valve (12 VDC, 20 PSIG)
1	562-864-06/7	Mouser	PMT/Power	AC EMI power filter/switch/fuse (6A rated)
1	H5784-06	Hamamatsu	PMT/Power	H5784-06 type photosensor module (185 nm to 650 nm)
1	BL-176302-U	Mouser	PMT/Power	±2V LCD voltmeter using 5V power supply
1	597-BB15-1.5	Mouser	PMT/Power	±12V Linear power supply
1	LE1182T-20K-1K-720B	TTE	Signal Proc.	20 kHz Elliptical low-pass filter (11 Pole, 1k impedance)
1	DT2801-A	Data Translation	Signal Proc.	DT2801-A 12-bit A/D board with ISA interface (8 MHz bus)
Small Electronic Components				
Qty.	Part #	Vendor	Stopped-Flow Component	Description
1	511-L78M05CV	Mouser	PMT/Power	LM7805
1		Mouser	PMT/Power	Fuse 120 VAC (1 A)
1	173-63102	Mouser	PMT/Power	Power cord
2	10YX034	Mouser	PMT/Power	3P4T Rotary dial switch
2	10TF115	Mouser	PMT/Power	DPDT switch
1	611-8168-202	Mouser	PMT/Power	SPDT ON-(ON) momentary push button switch
6	351-5111	Mouser	PMT/Power	T-1 3/4 LED (Red)
1	80-C330C104K1R	Mouser	PMT/Power	0.1 F ceramic disk capacitor
1	75-515D16V2200	Mouser	PMT/Power	2200 F electrolytic capacitor
10	291-330	Mouser	PMT/Power	330 resistor
10	291-510	Mouser	PMT/Power	500 resistor
10	291-1k	Mouser	PMT/Power	1 k resistor
10	291-2k	Mouser	PMT/Power	2 k resistor
10	291-5.1k	Mouser	PMT/Power	5 k resistor
1	311-1902-10k	Mouser	PMT/Power	10 k variable resistor potentiometer (single-turn)
100 ft.	602-3051-100-02	Mouser	PMT/Power	22 gauge hook-up wire (black)
100 ft.	602-3051-100-03	Mouser	PMT/Power	22 gauge hook-up wire (red)

Table A4.2 Complete parts list for stopped-flow modifications (continued) 254

Qty.	Part #	Vendor	Stopped- Flow Component	Description
100 ft.	602-3051-100-01	Mouser	PMT/Power	22 gauge hook-up wire (white)
100 ft.	602-3051-100-06	Mouser	PMT/Power	22 gauge hook-up wire (blue)
1	400-9043	Mouser	PMT/Power	Prototyping board (3"×1.85")
1	517-3425-6000	Mouser	PMT/Power	3M polarized 50 pin IDC conn.
5 ft.	517-3302/50FT	Mouser	PMT/Power	50 pin IDC cable
1	571-2075163	Mouser	PMT/Power	25 pin D-sub female (solder cup)
1	571-7479561	Mouser	PMT/Power	25 pin D-sub male/plug enclosure kit (or old printer cable)
100	571-1665060	Mouser	PMT/Power	Size 20 D-sub contacts/plugs crimp
100	571-1665040	Mouser	PMT/Power	Size 20 D-sub contacts /receptacles crimp
10 ft.	N/A	N/A	PMT/Power	10 to 15 conductor cable (just find an old printer cable and chop it up)
1	17RC007	Mouser	PMT/Power	5 pin DIN male connector in-line (pre-wired)
1	164-2513	Mouser	PMT/Power	5 pin DIN female connector in-line
1	174-2512	Mouser	PMT/Power	5 pin DIN male connector in-line
1	161-0005	Mouser	PMT/Power	5 pin DIN female panel mount
2	523-31-203-RFX	Mouser	PMT/Power	BNC receptacles panel mount 4-screw heavy duty
5	523-31-10-RFX	Mouser	PMT/Power	BNC receptacles panel mount
4	523-46650-93RFX	Mouser	PMT/Power	BNC terminators 93Ω
1	164-4205	Mouser	PMT/Power	Banana plug receptacle panel mount (red)
2	164-4201	Mouser	PMT/Power	Banana plug receptacle panel mount (black)
5	171-5462	Mouser	PMT/Power	Banana plug (red)
5	171-5461	Mouser	PMT/Power	Banana plug (black)
Hardware				
Qty.	Part #	Vendor	Stopped- Flow Component	Description
1		Oriel	Light Source	310 nm cut-on fluorescence filter
2			PMT/Power Signal Proc.	Aluminum Case 7"×11"×4"
10	561-L4.50	Mouser	PMT/Power	Nylon stand-off #4-40×½"
10	534-1893	Mouser	PMT/Power	Aluminum stand-off #4-40×½"
100	5721-440-3/8	Mouser	General	#4-40×3/8"
100	5721-440	Mouser	General	#4-40 nut zinc plated steel
1	532-507302B00	Mouser	PMT/Power	TO-220 package heat sink
2	450-2070	Mouser		Knob/dial

Figure A4.7 Signal quality of low-pass electronic filters. (Top) Fluorescence data were acquired at 50 μ s time intervals using either no filter (∇), a single 20 kHz (50 μ s), RC section (\circ), a 23 kHz, 11-pole elliptical function filter (\triangle), or both the RC and elliptical sections in serial (\square). Data traces are purposely offset for display only. *(Bottom)* The deviation histogram of each filtered data set from the mean using the same symbols as depicted above. Table A4.3 shows the quantitative comparison of the filters.

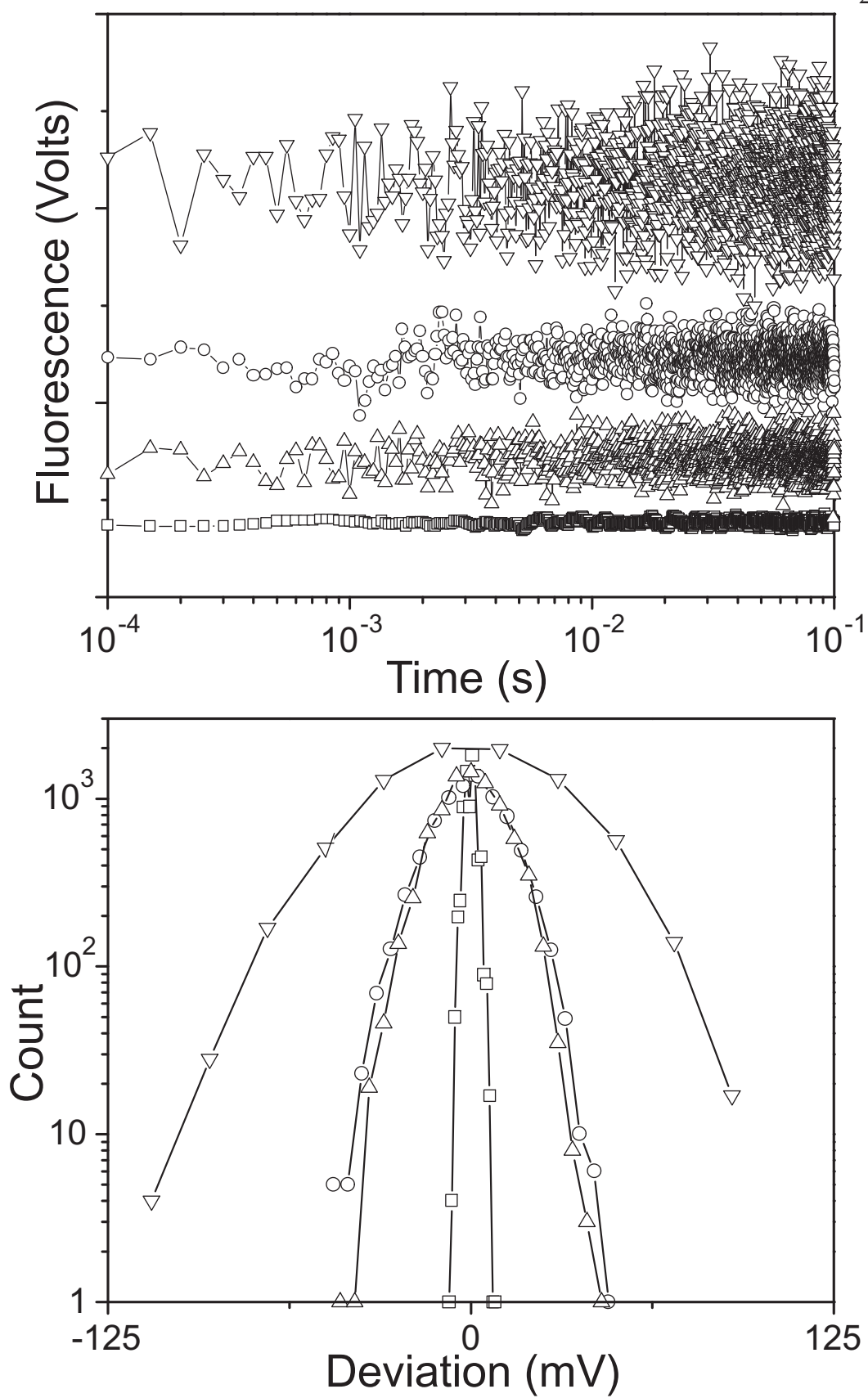


Table A4.3 Quantitative comparison of low-pass electronic signal filters

Description	Sum of Squares of Deviation from Mean	Root of Sum of Squares	Noise Reduction Ratio ^a
No filter	7.15	2.67	1.00
20 kHz RC section	1.30	1.14	2.35
23 kHz 11-pole elliptical function section	1.02	1.01	2.65
Both the RC and elliptical sections in serial	3.45×10^{-2}	0.19	14.39

^aRatio calculated from the root of the sum of the squares.

A4.2 Software

A4.2.1 Post-acquisition logarithmic filtering

When fitting multiphase kinetics data a 30-point per decade logarithmic sampling of the time course is very effective. Such a sampling reduces bias in over- or under-fitting certain portions of time course that may be less well represented in terms of amplitude.

A4.2.2 Microcal Origin software scripts

These scripts are intended to increase the productivity and throughput when determining the chevron plot for a protein at many denaturant concentrations. The scripts are designed to create Biokine filenames and MPS shot protocols while calculating the activation energies and denaturant concentrations on-the-fly. More advanced scripts can export SF4 files directly to the MPS software to reduce errors and time required in flipping between windows prior to each shot.

Table A4.4 Dataset names used by the Origin scripts

Dataset	Plot Designation	Description
Gdm	X	Denaturant concentration calculated here
RT	Y	RT natural log k_{obs} value calculated here
K	Disregard	k_{obs}
Syr1	Disregard	Volume delivered from syringe #1
Syr2	Disregard	Volume delivered from syringe #2
Syr3	Disregard	Volume delivered from syringe #3
Syr4	Disregard	Volume delivered from syringe #4
Shot	Disregard	Total volume of the shot
Filename	Disregard	Biokine denaturant indexed 8 character filename with extension
Cont1	Disregard	Denaturant concentration of syringe #1
Cont2	Disregard	Denaturant concentration of syringe #2
Cont3	Disregard	Denaturant concentration of syringe #3
Cont4	Disregard	Denaturant concentration of syringe #4
Temp	Disregard	Temperature of experiment stored in first position
Misc	Disregard	RT is calculated and stored in the first position

Table A4.5 Origin scripts

<p><i>Object name:</i> protocol</p> <p><i>Description:</i> Master command button which creates three and four syringe shot protocols and links to other scripts to calculate denaturant concentrations and create filenames.</p> <p><i>Event:</i> Mouse button-up</p>
<p><i>Command button text:</i></p> <p>\b(Syringe Protocol)</p> <p>\b(S1=\$(%H_cont1[1])M S2=\$(%H_cont2[1])M)</p> <p>\b(S3=\$(%H_cont3[1])M S4=\$(%H_cont4[1])M)</p>
<p><i>Script:</i></p> <pre> j1=0;j2=30;j3=11;j4=2; k1=0;k2=0;k3=12;k4=1; m1=0;m2=280;m3=10;m4=1; n1=0;n2=30;n3=11;n4=1; o1=0;o2=2;o3=250;o4=nn2;o5=1; q1=1; %O=(-10) (-9) (-8) (-7) (-6) (-5) (-4) (-3) (-2) (-1) 0 1 2 3 4 5 6 7 8 9 10; %M=(Three Syringe) (Four Syringe); getn (Syringe Mode) q1:M [Select Stopped Flow Protocol]; IF (q1==2) { getn (Syringe #1 [GdmHCl]) j1 (Starting Shot Volume) j2 (Shot increment) j3:0 [Syringe #1 Setup]; j2=abs(j2);j1=abs(j1);j3=(j3-11)*1; }; getn (Syringe #2 [GdmHCl]) k1 (Starting Shot Volume) k2 (Shot increment) k3:0 [Syringe #2 Setup]; k2=abs(k2);k1=abs(k1);k3=(k3-11)*1; getn (Syringe #3 [GdmHCl]) m1 (Starting Shot Volume) m2 (Shot increment) m3:0 [Syringe #3 Setup]; m2=abs(m2);m1=abs(m1);m3=(m3-11)*1; getn (Syringe #4 [GdmHCl]) n1 (Starting Shot Volume) n2 (Shot increment) n3:0 [Syringe #4 Setup]; n2=abs(n2);n1=abs(n1);n3=(n3-11)*1; define syrmath { loop(ii,0,500) { dummy=%4+(%5*ii); (%H_%1[ii+1])=dummy; IF (dummy<0) (%H_%1[ii+1])=0/0; </pre>

Table A4.4 Origin scripts (continued)

```

        (%H_%2[ii+1])=%3;
    };
};
IF (q1==1) {j2=0;j3=0;};
syrmath Syr1 cont1 j1 j2 j3;
syrmath Syr2 cont2 k1 k2 k3;
syrmath Syr3 cont3 m1 m2 m3;
syrmath Syr4 cont4 n1 n2 n3;
calcshots.run();
calcgdm.run();
creatfile.run();
(%H_TEMP[1])=0/0;
calcrt.run();

```

Object name: creatfile

Description: Creates biokine filenames using an integer calculated from the denaturant concentration, using the 8 + 3 character filename format. User may provide prefix and select suffix extension.

Event: Mouse button-up

Command button text:

Create
Filenames

Script:

```

%P=gdm;
i=1;
j=3;
%N=m.bkw .bkw (none);
%O=1 2 3 4;
getn
(Filename Prefix) %P
(Filename Suffix) i:N
(How many digits for Gdm) j:O
[Create Biokine Filenames];

%K=%H_Gdm;
%L=%H_Filename;
get %K -e numpoints;
For (ii=1; ii<=numpoints;ii++) {dummy=nint(10^(j-1)*(%K[ii]));
    If (i==1) {%S=%[N,#1]}
    If (i==2) {%S=%[N,#2]}
    If (i==3) {%S="" }
    (%L)[ii]=$%P$(dummy)%S;};

```

Object name: calcshots

Description: Short script to quickly add up total shot volume that is frequently called by other scripts.

Event: Mouse button-up

Command button text:

Calc Shot

Script: %K=%H_Syr1; %L=%H_Syr2; %M=%H_Syr3; %N=%H_Syr4; %O=%H_Shot;
 (%O)=((%K) + (%L) + (%M) + (%N));

Table A4.4 Origin scripts (continued)

<p><i>Object name:</i> calcgdm</p> <p><i>Description:</i> Short script to calculate denaturant concentration that is frequently called by other scripts.</p> <p><i>Event:</i> Mouse button-up</p>
<p><i>Command button text:</i></p> <p>Calc [GdmCl]</p>
<p><i>Script:</i></p> <pre>(%H_Gdm)=0/0; calcshots.run(); loop (ii,1,500) { ss1=%H_Syr1[ii]*%H_Cont1[ii]; ss2=%H_Syr2[ii]*%H_Cont2[ii]; ss3=%H_Syr3[ii]*%H_Cont3[ii]; ss4=%H_Syr4[ii]*%H_Cont4[ii]; stot=ss1+ss2+ss3+ss4; (%H_Gdm[ii])=stot/%H_Shot[ii]; };</pre>
<p><i>Object name:</i> calcgdm</p> <p><i>Description:</i> Short script to calculate activation energy of rate (k_{obs}) and autosave the origin project.</p> <p><i>Event:</i> Mouse button-up</p>
<p><i>Command button text:</i></p> <p>$\ln(RT) \ln(k \text{-(obs)})$ $(\\$(\%H_TEMP[1]) \backslash (176)C)$ $\ln(RT) = \\$(\%H_MISC)$</p>
<p><i>Script:</i></p> <pre>If ((%H_TEMP[1])==0/0) {getnumber (Degrees C) tmpc [Temperature Condition in Experiment]; (%H_TEMP[1])=tmpc; }; ELSE {tmpc=(%H_TEMP[1]);}; (%H_RT)=ln(%H_K)*0.001987192*(tmpc+273.17); (%H_MISC[1])=(tmpc+273.17)*0.001987192; save %X%G.opj;</pre>

Works Cited

- Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994). Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33, 10026-36.
- Agashe, V. R., Shastri, M. C. & Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature* 377, 754-7.
- Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. (2000). Stepwise formation of alpha-helices during cytochrome c folding. *Nature Struct. Biol.* 7, 514-520.
- Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230.
- Aune, K. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by GdmCl. *Biochemistry* 8, 4586-90.
- Avbelj, F., Luo, P. & Baldwin, R. L. (2000). Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl. Acad. Sci. U S A* 97, 10786-10791.
- Bai, Y., Englander, J. J., Mayne, L., Milne, J. S. & Englander, S. W. (1995a). Thermodynamic parameters from hydrogen exchange measurements. *Methods Enzymol.* 259, 344-56.
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* 17, 75-86.
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins* 20, 4-14.
- Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995b). Protein folding intermediates studied by native state hydrogen exchange. *Science* 269, 192-197.
- Baldwin, R. L. (1995). The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR* 5, 103-109.
- Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D. & Weaver, D. L. (1988). Diffusion-collision model for the folding kinetics of myoglobin. *Proteins Struct. Funct. Genet.* 4, 211-227.

- BenTal, N., Sitkoff, D., Topol, I. A., Yang, A. S., Burt, S. K. & Honig, B. (1997). Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *J. of Phys. Chem. B* 101, 450-457.
- Bhattacharyya, R. P. & Sosnick, T. R. (1999). Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil. *Biochemistry* 38, 2601-2609.
- Bhuyan, A. K. & Udgaonkar, J. B. (1999). Observation of multistate kinetics of folding and unfolding of barstar. *Biochemistry* 38, 9158-68.
- Bhuyan, A. K. & Udgaonkar, J. B. (2001). Folding of horse cytochrome c in the reduced state. *J. Mol. Biol.* 312, 1135-60.
- Bieri, O. & Kiefhaber, T. (2001). Origin of apparent fast and non-exponential kinetics of lysozyme folding measured in pulsed hydrogen exchange experiments. *J. Mol. Biol.* 310, 919-35.
- Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA* 96, 9597-601.
- Bolhuis, P. G., Dellago, C. & Chandler, D. (2000). Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. USA* 97, 5877-82.
- Bowers, P. M. & Klevit, R. E. (1996). Hydrogen bonding and equilibrium isotope enrichment in histidine- containing proteins. *Nature Struct. Biol.* 3, 522-31.
- Bowers, P. M. & Klevit, R. E. (2000). Hydrogen bond geometry and H-2/H-1 fractionation in proteins. *J. Am. Chem. Soc.* 122, 1030-1033.
- Branden, C. & Tooze, J. (1991). *Introduction to Protein Structure*, Garland Publishing, Inc., New York, NY.
- Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to cis trans isomerism of proline residues. *Biochemistry* 14, 4953-4963.
- Briggs, M. S. & Roder, H. (1992). Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. Natl. Acad. Sci. USA* 89, 2017-21.
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21, 167-195.

- Bulaj, G. & Goldenberg, D. P. (2001). Phi-values for BPTI folding intermediates and implications for transition state analysis. *Nature Struct. Biol.* 8, 326-330.
- Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. (1997). The energy landscape of a fast-folding protein mapped by Ala-->Gly substitutions. *Nature Struct. Biol.* 4, 305-310.
- Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W. & Oas, T. G. (1996). Microsecond protein folding through a compact transition state. *J. Mol. Biol.* 263, 311-322.
- Burton, R. E., Myers, J. K. & Oas, T. G. (1998). Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* 37, 5337-43.
- Bushnell, G. W., Louie, G. V. & Brayer, G. D. (1990). High-resolution three dimensional structure of horse heart cytochrome c. *J. Mol. Biol.* 213, 585-595.
- Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. (1990). Detection and characterization of a folding intermediate in barnase by NMR. *Nature* 346, 488-490.
- Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. & Radford, S. E. (2001). Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nature Struct. Biol.* 8, 68-72.
- Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Science*. 8, 2251-7.
- Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature Struct. Biol.* 3, 782-787.
- Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofrichter, J. (1997). Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. USA* 94, 1779-1784.
- Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. & Doniach, S. (1998). Kinetics of lysozyme refolding: structural characterization of a non- specifically collapsed state using time-resolved X-ray scattering. *J. Mol. Biol.* 276, 225-37.
- Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* 6, 1005-9.
- Chu, R. A., Takei, J., Barchi, J. J., Jr. & Bai, Y. (1999). Relationship between the native-state hydrogen exchange and the folding pathways of barnase. *Biochemistry* 38, 14119-24.

- Clarke, J. & Fersht, A. R. (1996). An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Fold. Des.* 1, 243-54.
- Cleland, W. W. (1995). Isotope effects: determination of enzyme transition state structure. *Methods Enzymol.* 249, 341-373.
- Colon, W., Elöve, G., Wakem, L. P., Sherman, F. & Roder, H. (1996). Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry* 35, 5538-49.
- Connelly, G. P., Bai, Y., Jeng, M.-F., Mayne, L. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins* 17, 87-92.
- Crane, J. C., Koepf, E. K., Kelly, J. W. & Gruebele, M. (2000). Mapping the transition state of the WW domain beta-sheet. *J. Mol. Biol.* 298, 283-92.
- Creighton, T. E. (1977). Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J. Mol. Biol.* 113, 275-293.
- Creighton, T. E. (1978). Experimental studies of protein folding and unfolding. *Prog. Biophys. Mol. Biol.* 33, 231-297.
- Creighton, T. E. (1985). The problem of how and why proteins adopt folded conformations. *J. Phys. Chem.* 89, 2452-2459.
- Creighton, T. E. (1997). Protein folding: does diffusion determine the folding rate? *Curr. Biol.* 7, R380-3.
- Debe, D. A. & Goddard, W. A., 3rd. (1999). First principles prediction of protein folding rates. *J. Mol. Biol.* 294, 619-25.
- Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.
- Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4, 19.
- Dill, K. A., Fiebig, K., M. & Chan, H. S. (1993). Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. USA* 90, 1942-1946.
- Durr, E., Jelesarov, I. & Bosshard, H. R. (1999). Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices. *Biochemistry* 38, 870-880.
- Eaton, W. A., Munoz, V., Thompson, P. A., Chan, C.-K. & Hofrichter, J. (1997). Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7, 10-14.

- Edison, A. S., Weinhold, F. & Markley, J. L. (1995). Theoretical studies of protium/deuterium fractionation factors and cooperative hydrogen bonding in peptides. *J. Am. Chem. Soc.* 117, 9619-9624.
- Eigenbrot, C., Randal, M. & Kossiakoff, A. A. (1990). Structural effects induced by removal of a disulfide-bridge: the X-ray structure of the C30A/C51A mutant of basic pancreatic trypsin inhibitor at 1.6 Å. *Protein Eng.* 3, 591-8.
- Ellis, R. J. (2001). Molecular chaperones: inside and outside the Anfinsen cage. *Curr. Biol.* 11, R1038-40.
- Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* 29, 213-38.
- Englander, S. W. & Mayne, L. (1992). Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* 21, 243-65.
- Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X. & Bai, Y. (1998). Fast and Slow Folding in Cytochrome C. *Accts. of Chem. Res.* 31, 737-744.
- Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* 286, 1597-1608.
- Fernandez, A., Colubri, A. & Appignane, G. (2001). Semiempirical prediction of protein folds. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 64, 021901.
- Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA* 92, 10869-73.
- Fersht, A. R. (2000a). A kinetically significant intermediate in the folding of barnase. *Proc. Natl. Acad. Sci. USA* 97, 14121-6.
- Fersht, A. R. (2000b). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA* 97, 1525-9.
- Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl. Acad. Sci. USA* 91, 10426-9.
- Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224, 771-782.

- Fischer, G. & Bang, H. (1985). The refolding of urea-denatured ribonuclease A is catalyzed by peptidyl-prolyl cis-trans isomerase. *Biochim. Biophys. Acta* 828, 39-42.
- Garel, J. R. & Baldwin, R. L. (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Natl. Acad. Sci. U S A* 70, 3347-51.
- Garel, J.-R., Nall, B. T. & Baldwin, R. L. (1976). Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. *Proc. Natl. Acad. Sci. USA* 93, 1853-1857.
- Ghadiri, M. & Choi, C. (1990). Secondary structure nucleation in peptides - Transition-metal ion stabilized alpha-helices. *J. Amer. Chem. Soc.* 112, 1630-1632.
- Gladwin, S. T. & Evans, P. A. (1996). Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling. *Fold. Des.* 1, 407-17.
- Goldenberg, D. P. (1999). Finding the right fold. *Nature Struct. Biol.* 6, 987-990.
- Gorski, S. A., Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2001). Acidic conditions stabilise intermediates populated during the folding of im7 and im9. *J. Mol. Biol.* 312, 849-63.
- Grantcharova, V., Alm, E. J., Baker, D. & Horwich, A. L. (2001). Mechanisms of protein folding. *Curr. Opin. Struct. Biol.* 11, 70-82.
- Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry* 36, 15685-15692.
- Grantcharova, V. P. & Baker, D. (2001). Circularization changes the folding transition state of the src SH3 domain. *J. Mol. Biol.* 306, 555-563.
- Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc. Natl. Acad. Sci. U S A* 97, 7084-7089.
- Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nature Struct. Biol.* 5, 714-720.
- Grathwohl, C. & Wuthrich, K. (1976a). NMR studies of the molecular conformations in the linear oligopeptides H-(L-Ala)_n-L-Pro-OH. *Biopolymers* 15, 2043-57.
- Grathwohl, C. & Wuthrich, K. (1976b). The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers* 15, 2025-41.
- Gruebele, M. & Wolynes, P. G. (1998). Satisfying turns in folding transitions. *Nature Struct. Biol.* 5, 662-5.

- Gu, H., Kim, D. & Baker, D. (1997). Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. *J. Mol. Biol.* 274, 588-96.
- Guerois, R. & Serrano, L. (2000). The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* 304, 967-82.
- Guo, Z. & Thirumalai, D. (1997). The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus. *Fold. Des.* 2, 377-91.
- Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995). Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* 34, 3066-76.
- Hagen, S. J. & Eaton, W. A. (2000). Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.* 297, 781-9.
- Hartl, F. U. & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852-8.
- Hennecke, J., Sebbel, P. & Glockshuber, R. (1999). Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J. Mol. Biol.* 286, 1197-215.
- Honeycutt, J. D. & Thirumalai, D. (1992). The nature of folded states of globular proteins. *Biopolymers* 32, 695-709.
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216, 1031-44.
- Hosszu, L. L., Craven, C. J., Parker, M. J., Lorch, M., Spencer, J., Clarke, A. R. & Waltho, J. P. (1997). Structure of a kinetic protein folding intermediate by equilibrium amide exchange. *Nature Struct. Biol.* 4, 801-4.
- Huang, G. S. & Oas, T. G. (1995). Submillisecond folding of monomeric lambda repressor. *Proc. Natl. Acad. Sci. USA* 92, 6878-82.
- Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Adv. Protein Chem.* 21, 287-386.
- Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I. & Sanchez-Ruiz, J. M. (1999). Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* 38, 8138-49.
- Ibarra-Molero, B., Makhatadze, G. I. & Matthews, C. R. (2001). Mapping the energy surface for the folding reaction of the coiled-coil peptide GCN4-p1. *Biochemistry* 40, 719-31.

- Itzhaki, L. S. & Evans, P. A. (1996). Solvent isotope effects on the refolding kinetics of hen egg-white lysozyme. *Protein Science*. 5, 140-146.
- Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat Gay, G. & Fersht, A. R. (1995a). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* 254, 289-304.
- Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995b). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254, 260-288.
- Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* 3, R81-91.
- Jackson, S. E. & Fersht, A. R. (1991a). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428-10435.
- Jackson, S. E. & Fersht, A. R. (1991b). Folding of chymotrypsin inhibitor 2. 2. Influence of proline. *Biochemistry* 30, 10436-10443.
- Jacob, M., Geeves, M., Holtermann, G. & Schmid, F. X. (1999). Diffusional barrier crossing in a two-state protein folding reaction. *Nature Struct. Biol.* 6, 923-6.
- Jacob, M., Schindler, T., Balbach, J. & Schmid, F. X. (1997). Diffusion control in an elementary protein folding reaction. *Proc. Natl. Acad. Sci. USA* 94, 5622-7.
- Jacob, M. & Schmid, F. X. (1999). Protein folding as a diffusional process. *Biochemistry* 38, 13773-9.
- Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. (2001). The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* 311, 373-93.
- Jancso, G. V. H., W. Alexander. (1974). Condensed phase isotope effects. *Chem. Rev.* 74, 689-719.
- Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Science* 3, 650-68.
- Kauzmann, E. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14, 1-64.
- Kelley, R. F. & Richards, F. M. (1987). Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* 26, 6765-74.
- Kentsis, A. & Sosnick, T. R. (1998). Trifluoroethanol promotes helix formation by destabilizing backbone exposure: Desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding. *Biochemistry* 37, 14613-14622.

- Khare, D., Alexander, P. & Orban, J. (1999). Hydrogen bonding and equilibrium protium-deuterium fractionation factors in the immunoglobulin G binding domain of protein G. *Biochemistry* 38, 3918-25.
- Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 32, 7054-7063.
- Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a 3-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* 3, 193-205.
- Kiefhaber, T., Grunert, H. P., Hahn, U. & Schmid, F. X. (1990). Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry* 29, 6475-80.
- Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M. & Baker, D. (1998). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* 284, 807-15.
- Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51, 459-89.
- Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59, 631-660.
- Krantz, B. A., Moran, L. B., Kentsis, A. & Sosnick, T. R. (2000). D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nature Struct. Biol.* 7, 62-71.
- Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* 39, 11696-701.
- Krantz, B. A. & Sosnick, T. R. (2001). Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nature Struct. Biol.* 8, 1042-1047.
- Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nature Struct. Biol.* 9, 458-63.
- Kuszewski, J., Clore, G. M. & Gronenborn, A. M. (1994). Fast folding of a prototypic polypeptide: The immunoglobulin binding domain of streptococcal protein G. *Protein Science* 3, 1945-1952.
- Lang, K., Schmid, F. X. & Fischer, G. (1987). Catalysis of protein folding by prolyl isomerase. *Nature* 329, 268-70.

- Lazaridis, T., Archontis, G. & Karplus, M. (1995). Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv. Protein Chem.* 47, 231-306.
- Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* 89, 8721-8725.
- Levinthal, C. (1968). Are there pathways for protein folding. *J. Chim. Phys.* 65, 44-45.
- Levinthal, C. (1969). *Mossbauer Spectroscopy in Biological Systems (proceedings of a meeting held at Allerton House, Monticello, Ill.)*.
- Levitt, M. (1981). Effect of proline residues on protein folding. *J. Mol. Biol.* 145, 251-63.
- LiWang, A. C. & Bax, A. (1996). Equilibrium protium/deuterium fractionation of backbone amides in U-13C/15N labeled human ubiquitin by triple resonance NMR. *J. Am. Chem. Soc.* 118, 12864-12865.
- Loftus, D., Gbenle, G. O., Kim, P. S. & Baldwin, R. L. (1986). Effects of denaturants on amide proton exchange rates. A test for structure in protein fragments and folding intermediates. *Biochemistry* 25, 1428-1436.
- Loh, S. N. & Markley, J. L. (1993). *Measurement of amide hydrogen D/H fractionation factors in proteins by NMR spectroscopy*. Techniques in protein chemistry IV (Angelletti, R., Ed.), Academic Press, San Diego.
- Loh, S. N. & Markley, J. L. (1994). Hydrogen bonding in proteins as studied by amide hydrogen D/H fractionation factors: application to staphylococcal nuclease. *Biochemistry* 33, 1029-36.
- Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. (2001). Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Science* 10, 1343-52.
- Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* 38, 16419-23.
- Loladze, V. V. & Makhatadze, G. I. (2002). Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. *Protein Science* 11, 174-7.
- Lorch, M., Mason, J. M., Clarke, A. R. & Parker, M. J. (1999). Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. *Biochemistry* 38, 1377-85.
- Lorch, M., Mason, J. M., Sessions, R. B. & Clarke, A. R. (2000). Effects of mutations on the thermodynamics of a protein folding reaction: implications for the mechanism of formation of the intermediate and transition states. *Biochemistry* 39, 3480-5.

- Lumry, R., Biltonen, R. & Brandts, J. F. (1966). Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* 4, 917-44.
- Makhatadze, G. I. (1998). Heat capacities of amino acids, peptides and proteins. *Biophys. Chem.* 71, 133-56.
- Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). Solvent isotope effect and protein stability. *Nature Struct. Biol.* 2, 852-855.
- Makhatadze, G. I., Lopez, M. M., Richardson, J. M., 3rd & Thomas, S. T. (1998). Anion binding to the ubiquitin molecule. *Protein Science* 7, 689-97.
- Makhatadze, G. I. & Privalov, P. L. (1990). Heat capacity of proteins I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. *J. Mol. Biol.* 213, 375-384.
- Martinez, J. C., Pisabarro, M. T. & Serrano, L. (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.* 5, 721-9.
- Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* 6, 1010-6.
- Martinez, J. C., Viguera, A. R., Berisio, R., Wilmanns, M., Mateo, P. L., Filimonov, V. V. & Serrano, L. (1999). Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: discerning the reasons for rapid folding in proteins. *Biochemistry* 38, 549-59.
- Matheson, R. R. & Scheraga, H. A. (1978). A method for predicting nucleation sites for protein folding based upon hydrophobic contacts. *Macromolecules* 11, 814-829.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* 346, 440-5.
- Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry* 34, 13656-62.
- Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M. & Fersht, A. R. (1992). The folding of an enzyme. V. H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J. Mol. Biol.* 224, 837-45.
- Matthews, C. R. (1987). Effects of point mutations on the folding of globular proteins. *Methods Enzymol.* 154, 498-511.
- Mayne, L. & Englander, S. W. (2000). Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Science* 9, 1873-7.

- McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Struct. Biol.* 7, 669-673.
- Milne, J. S., Mayne, L., Roder, H., Wand, A. J. & Englander, S. W. (1998). Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Science* 7, 739-45.
- Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A. & Sosnick, T. R. (1999). Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc. Natl. Acad. Sci. USA* 96, 10699-10704.
- Moult, J. & Unger, R. (1991). An analysis of protein folding pathways. *Biochemistry* 30, 3816-24.
- Munoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* 1, 399-409.
- Munoz, V. & Eaton, W. A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA* 96, 11311-6.
- Munoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998). A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. USA* 95, 5872-9.
- Munoz, V. & Serrano, L. (1997). Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41, 495-509.
- Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature* 390, 196-9.
- Myers, J. K. & Oas, T. G. (1999). Reinterpretation of GCN4-p1 Folding Kinetics: Partial Helix Formation Precedes Dimerization in Coiled Coil Folding. *J. Mol. Biol.* 289, 205-209.
- Myers, J. K. & Oas, T. G. (2001). Preorganized secondary structure as an important determinant of fast protein folding. *Nature Struct. Biol.* 8, 552-8.
- Myers, J. K. & Pace, C. N. (1996). Hydrogen bonding stabilizes globular proteins. *Biophysical J.* 71, 2033-2039.
- Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science* 4, 2138-48.
- Naderi, H. M., Thomason, J. F., Borgias, B. A., Anderson, S., James, T. L. & Kuntz, I. D. (1991). ¹H NMR assignments and three-dimensional structure of Ala14/Ala38 bovine pancreatic trypsin inhibitor based on two-dimensional NMR and distance geometry. In *Conformations and Forces in Protein Folding* (Nall, B. T. & Dill, K. A., eds.), pp. 86-114. Am. Assoc. Adv. Sci., Washington D.C.
- Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nature Struct. Biol.* 8, 602-605.

- Northrop, D. B. (1981). The expression of isotope effects on enzyme-catalyzed reactions. *Annu. Rev. Biochem.* 50, 103-131.
- Nymeyer, H., Garcia, A. E. & Onuchic, J. N. (1998). Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* 95, 5921-8.
- Oliveberg, M., Tan, Y. J., Silow, M. & Fersht, A. R. (1998). The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* 277, 933-43.
- Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. (1996). Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* 1, 441-50.
- Otzen, D. E. & Fersht, A. R. (1998). Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. *Biochemistry* 37, 8139-46.
- Otzen, D. E., Kristensen, O., Proctor, M. & Oliveberg, M. (1999). Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* 38, 6499-511.
- Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* 10, 75-83.
- Park, S. H., Shastry, M. C. & Roder, H. (1999). Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nature Struct. Biol.* 6, 943-7.
- Parker, M. J. & Clarke, A. R. (1997). Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry* 36, 5786-5794.
- Parker, M. J., Dempsey, C. E., Lorch, M. & Clarke, A. R. (1997). Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. *Biochemistry* 36, 13396-405.
- Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst-phase reactions explored. *J. Mol. Biol.* 293, 1195-210.
- Parker, M. J. & Marqusee, S. (2000). A statistical appraisal of native state hydrogen exchange data: evidence for a burst-phase continuum? *J. Mol. Biol.* 300, 1361-75.
- Parker, M. J. & Marqusee, S. (2001). A Kinetic Folding Intermediate Probed by Native State Hydrogen Exchange. *J. Mol. Biol.* 305, 593-602.
- Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* 313, 343-57.

- Plaxco, K. W. & Baker, D. (1998). Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl. Acad. Sci. USA* 95, 13591-6.
- Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985-994.
- Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker, D. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* 39, 11177-11183.
- Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999). Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA* 96, 10115-7.
- Privalov, P. L. & Makhatadze, G. I. (1990). Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects. *J. Mol. Biol.* 213, 385-91.
- Ptitsyn, O. B. (1995). How the molten globule became. *Trends in Biochemical Sciences* 20, 376-9.
- Ptitsyn, O. B. (1998). Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes? *J. Mol. Biol.* 278, 655-66.
- Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998). The burst-phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nature Struct. Biol.* 5, 882-4.
- Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358, 302-307.
- Ramdas, L. & Nall, B. T. (1986). Folding/unfolding kinetics of mutant forms of iso-1-cytochrome c with replacement of proline-71. *Biochemistry* 25, 6959-64.
- Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nature Struct. Biol.* 6, 825-31.
- Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct. Biol.* 4, 298-304.

- Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* 6, 1016-1024.
- Ristow, S. S. & Wetlaufer, D. B. (1973). Evidence for nucleation in the folding of reduced hen egg lysozyme. *Biochem. Biophys. Res. Commun.* 50, 544-50.
- Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* 7, 15-28.
- Roder, H., Elöve, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR. *Nature* 335, 700-704.
- Ross, P. D. & Rekharsky, M. V. (1996). Thermodynamics of hydrogen bond and hydrophobic interactions in cyclodextrin complexes. *Biophysical J.* 71, 2144-54.
- Rumbley, J., Hoang, L., Mayne, L. & Englander, S. W. (2001). An amino acid code for protein folding. *Proc. Natl. Acad. Sci. USA* 98, 105-12.
- Sabelko, J., Ervin, J. & Gruebele, M. (1999). Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. USA* 96, 6031-6.
- Sancho, J., Meiering, E. M. & Fersht, A. R. (1991). Mapping transition states of protein unfolding by protein engineering of ligand-binding sites. *J. Mol. Biol.* 221, 1007-14.
- Sauder, J. M. & Roder, H. (1998). Amide protection in an early folding intermediate of cytochrome c. *Fold. Des.* 3, 293-301.
- Scheiner, S. (2000). Calculation of isotope effects from first principles. *Biochim. Biophys. Acta* 1458, 28-42.
- Schonbrunner, N., Pappenberger, G., Scharf, M., Engels, J. & Kiefhaber, T. (1997). Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for beta- sheet formation. *Biochemistry* 36, 9057-65.
- Schowen, K. B. & Schowen, R. L. (1982). Solvent isotope effects of enzyme systems. *Methods Enzymol.* 87, 551-606.
- Segawa, S. & Sugihara, M. (1984). Characterization of the transition state of lysozyme unfolding. II. Effects of the intrachain crosslinking and the inhibitor binding on the transition state. *Biopolymers* 23, 2489-98.
- Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* 29, 9343-52.

- Shakhnovich, E. I. (1998). Folding nucleus: specific or multiple? Insights from lattice models and experiments. *Fold. Des.* 3, R108-11.
- Shastry, M. C., Luck, S. D. & Roder, H. (1998). A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. *Biophysical J.* 74, 2714-21.
- Shastry, M. C. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.* 5, 385-92.
- Shi, Z., Krantz, B. A., Kallenbach, N. & Sosnick, T. R. (2002). Contribution of Hydrogen Bonding to Protein Stability Estimated from Isotope Effects. *Biochemistry* 41, 2120-2129.
- Silow, M. & Oliveberg, M. (1997). High-energy channeling in protein folding. *Biochemistry* 36, 7633-7.
- Silow, M., Tan, Y. J., Fersht, A. R. & Oliveberg, M. (1999). Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry* 38, 13006-12.
- Socchi, N. D., Onuchic, J. N. & Wolynes, P. G. (1998). Protein folding mechanisms and the multidimensional folding funnel. *Proteins* 32, 136-58.
- Sosnick, T. R. & Englander, S. W. (1996). What limits protein folding. In *Dynamics and the problem of recognition in biological macromolecules* (Jardetzky, O. & Lefevre, J.-F., eds.), Vol. 288, pp. 65-71. Plenum Press, NY.
- Sosnick, T. R., Jackson, S., Wilk, R. M., Englander, S. W. & DeGrado, W. F. (1996a). The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins* 24, 427-432.
- Sosnick, T. R., Mayne, L. & Englander, S. W. (1996b). Molecular collapse: The rate-limiting step in two-state cytochrome c folding. *Proteins* 24, 413-426.
- Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* 1, 149-156.
- Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1995). *Peptide and Protein Folding Workshop, Philadelphia, PA.*
- Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. USA* 94, 8545-50.

- Srivastava, A. K. & Sauer, R. T. (2000). Evidence for partial secondary structure formation in the transition state for arc repressor refolding and dimerization. *Biochemistry* 39, 8308-8314.
- Staley, J. P. & Kim, P. S. (1992). Complete folding of BPTI with only a single disulfide bond. *Proc. Natl. Acad. Sci. USA* 89, 1519-23.
- Stassinopoulou, C. I., Wagner, G. & Wuthrich, K. (1984). Two-dimensional ^1H NMR of two chemically modified analogs of the basic pancreatic trypsin inhibitor. Sequence-specific resonance assignments and sequence location of conformation changes relative to the native protein. *Eur. J. Biochem.* 145, 423-30.
- States, D. J., Dobson, C. M. & Karplus, M. (1984). A new two-disulphide intermediate in the refolding of reduced bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 174, 411-8.
- Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M., Serrano, L., Dobson, C. M. & Ramponi, G. (2000). Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* 300, 633-647.
- Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S. & Rousseau, D. L. (1997). Folding of cytochrome c initiated by submillisecond mixing. *Nature Struct. Biol.* 4, 44-50.
- Takei, J., Chu, R. A. & Bai, Y. (2000). Absence of stable intermediates on the folding pathway of barnase. *Proc. Natl. Acad. Sci. U S A* 97, 10796-801.
- Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* 23, 121-282.
- Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* 24, 1-95.
- Tanford, C. (1980). *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, John Wiley & Sons, New York.
- Thompson, P. A., Eaton, W. A. & Hofrichter, J. (1997). Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* 36, 9200-10.
- Thirumalai, D. & Guo, Z. (1995). Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange experiments. *Biopolymers* 35, 137-140.
- Thirumalai, D. & Klimov, D. K. (1998). Fishing for folding nuclei in lattice models and proteins. *Fold. Des.* 3, R112-8.
- Travaglini-Allocatelli, C., Cutruzzola, F., Bigotti, M. G., Staniforth, R. A. & Brunori, M. (1999). Folding mechanism of *Pseudomonas aeruginosa* cytochrome c551: role of electrostatic interactions on the hydrophobic collapse and transition state properties. *J. Mol. Biol.* 289, 1459-67.
- Viguera, A. R. & Serrano, L. (1997). Loop length, intramolecular diffusion and protein folding. *Nature Struct. Biol.* 4, 939-46.

- Viguera, A. R., Serrano, L. & Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nature Struct. Biol.* 3, 874-80.
- Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., Vierstra, R. D., Hatfield, P. M. & Cook, W. J. (1987). Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin. *J. Biol. Chem.* 262, 6396-9.
- Wade, D. (1999). Deuterium isotope effects on noncovalent interactions between molecules. *Chem. Biol. Interact.* 117, 191-217.
- Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. USA* 93, 2629-2634.
- Warshel, A. & Russell, S. T. (1984). Calculations of electrostatic interactions in biological systems and in solutions. *Quarterly Reviews of Biophysics* 17, 283-422.
- Weissman, J. S. & Kim, P. S. (1991). Reexamination of the folding of BPTI: Predominance of native intermediates. *Science* 253, 1386-1393.
- Weissman, J. S. & Kim, P. S. (1992a). Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U S A* 89, 9900-4.
- Weissman, J. S. & Kim, P. S. (1992b). The pro region of BPTI facilitates folding. *Cell* 71, 841-51.
- Weissman, J. S. & Kim, P. S. (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature* 365, 185-8.
- Weissman, J. S. & Kim, P. S. (1995). A kinetic explanation for the rearrangement pathway of BPTI folding. *Nature Struct. Biol.* 2, 1123-30.
- Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* 70, 697-701.
- White, T. B., Berget, P. B. & Nall, B. T. (1987). Changes in conformation and slow refolding kinetics in mutant iso-2-cytochrome c with replacement of a conserved proline residue. *Biochemistry* 26, 4358-66.
- Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270, 294-304.
- Wuthrich, K. & Grathwohl, C. (1974). A novel approach for studies of the molecular conformations in flexible polypeptides. *FEBS Lett.* 43, 337-40.
- Yang, A. S. & Honig, B. (1995). Free energy determinants of secondary structure formation: I. alpha- Helices. *J. Mol. Biol.* 252, 351-365.
- Yeh, S. R. & Rousseau, D. L. (1998). Folding intermediates in cytochrome c. *Nature Struct. Biol.* 5, 222-8.
- Yeh, S.-R., Takahashi, S., Fan, B. & Rousseau, D. L. (1997). Ligand exchange during cytochrome c folding. *Nature Struct. Biol.* 4, 51-56.

- Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D. & Chu, S. (2000). A single-molecule study of RNA catalysis and folding. *Science* 288, 2048-51.
- Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* 296, 1105-16.